AD_____

Award Number: DAMD17-97-1-7118

TITLE: Few-View Tomographic Reconstruction of Technetium-
99m-sestamibi Distribution for the Detection and
Differentiation of Breast Lesions

PRINCIPAL INVESTIGATOR: Patrick La Rivière

Chin-Tu Chen

CONTRACTING ORGANIZATION: The University of Chicago
Chicago, Illinois   60637

REPORT DATE: August 2000

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are
those of the author(s) and should not be construed as an official
Department of the Army position, policy or decision unless so
designated by other documentation.

20010419 091

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 2000	Final (15 Jul 97 - 14 Jul 00)

4. TITLE AND SUBTITLE
Few-View Tomographic Reconstruction of Technetium-99m-sestamibi Distribution for the Detection and Differentiation of Breast Lesions

5. FUNDING NUMBERS
DAMD17-97-1-7118

6. AUTHOR(S)
Patrick La Rivière

Chin-Tu Chen

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of Chicago
Chicago, Illinois 60637

E-MAIL:
pjlarivi@midway.uchicago.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

Scintimammography (SMM) is a nuclear medicine test with the potential to provide relatively low-cost, minimally invasive differentiation of breast abnormalities detected by physical examination or mammography. While the most widely used clinical protocol involves acquiring one or two planar views, occasionally supplemented by conventional SPECT, we have established that a dedicated breast SPECT geometry, in which the camera revolves around the breast alone, would provide superior lesion detectability to these other two approaches. Because the time required by tomographic studies to acquire data useful to popular reconstruction algorithms might be excessive for SPECT SMM, we have worked to develop reconstruction algorithms that can generate diagnostically useful SMM images from a smaller number of views than is usually used. In particular, we have developed a technique in which the few-view sinogram is first smoothed using a spline-based, Bayesian technique and then additional views are interpolated using periodic spline interpolation. The spline interpolation approach was chosen after extensive investigation of the accuracy and noise properties of various periodic interpolation approaches. We find that with use of this technique, diagnostically meaningful dedicated SPECT SMM images can be reconstructed from as few as 15 projection views.

14. SUBJECT TERMS
Breast Cancer, scintimammography, tomography, SPECT, Tc-99m-sestambi, interpolation, smoothing

15. NUMBER OF PAGES
125

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

1 Introduction

The broad objective of the proposed research was to develop, implement, and evaluate methods for the reconstruction of dedicated single-photon emission computed tomography (SPECT) scintimammography (SMM) images from a relatively small number of projection views. SMM is a nuclear-medicine test with the potential to provide relatively low-cost, minimally invasive differentiation of breast abnormalities identified by physical examination or mammography. It relies on the preferential uptake of Tc-99m-sestamibi or other radionuclides in breast malignancies as compared to normal breast tissue or benign abnormalities. This focal uptake can be imaged in a number of ways, though the most widely used clinical protocol involves acquiring one or two planar views while the patient lies prone on a specially designed table. However, preliminary experimental work, verified theoretically in our own work, has suggested that a dedicated breast SPECT geometry, in which a small camera revolves around a dependent breast, would provide better lesion detectability than do the planar or conventional SPECT geometries. The drawback of this approach would be the relatively long time needed to acquire the number of projection views traditionally used for SPECT image reconstruction. The aim of this research, then, was to reduce imaging time in dedicated breast SPECT by developing algorithms that allow images to be reconstructed from a smaller number of projection views than is conventionally used while maintaining diagnostically useful image quality. Whereas the original strategy was to develop fully iterative reconstruction techniques incorporating prior assumptions about the structures being imaged, these assumptions proved difficult to quantify, the resulting algorithms computationally intensive, and the reconstructed images relatively poor. Moreover, many unresolved issues regarding regularization and stopping rules plague these iterative approaches. We have thus pursued an alternative strategy involving sinogram preprocessing, in which each projection view is first smoothed using Fourier or spline-based, Bayesian techniques and then additional projection views are interpolated, again using Fourier or spline-based techniques, prior to reconstruction by filtered backprojection (FBP). The physical characteristics of the algorithms, such as resolution and noise properties, were to be characterized through the use of simulated data, and the methods then applied to real phantom and clinical patient data.

2 Body

While the unexpected findings that always accompany a research project lead us in some directions that were not anticipated in the original proposal, all of the tasks detailed in the original statement of work were accomplished or addressed in some form, with the exception of the last two (tasks 9 and 10), pertaining to technical objective 5. This objective involved testing the algorithms in a clinical setting, with the cooperation of Dr. Patrick Peller of Lutheran General Hospital in Chicago, who had proposed to develop a dedicated breast imaging miniature gamma camera in collaboration with the University of Chicago. This project, which was independent of the one being discussed in this report, did not come to fruition, and thus no clinical SPECT SMM yet exists on which to test the algorithms. To compensate for this, more attention was paid to the analytic characterization of the algorithms' performance.

The remainder of this section will describe the research accomplishments associated with each task outlined in the original statement of work.

Technical Objective 1

To develop algebraic reconstruction technique-based algorithms for the reconstruction of MIBI distribution in breasts from few views, including some algorithms that reconstruct a purely binary reconstruction image.

2.1 Task 1—Implementing ART

Finish implementing standard algebraic reconstruction technique algorithm.

4

We succeeded in implementing the class of algorithms known as algebraic reconstruction techniques (ART) for few-view reconstruction. Unlike the clinically standard filtered backprojection (FBP) algorithm, ART algorithms make no implicit assumptions about the completeness of the data but rather seek to produce iteratively a solution that best matches the data that is available, according to some precise definition of "best." Thus, because ART is an explicitly discrete algorithm rather than a discretization of an algorithm based upon a continuous analysis, our hypothesis was that the ART algorithms would fare better than FBP in reconstructions from limited numbers of views. In fact, this hypothesis was not entirely supported by our findings. Even in the absence of noise, the quality of the reconstructed images as compared to FBP was not sufficiently high to justify the large computational burden of the approach. In the presence of noise, images reconstructed by the ART approach were really quite poor, even compared to ramp-filtered FBP. These same general trends held for reconstructions from standard and small numbers of views. The methods and these findings are now discussed in greater detail.

2.1.1 Methods—ART

In their most basic form, ART algorithms begin with an initial "guess" of the image, often just a uniform image, then project this estimate and compare the result to the measured data. The resulting differences are used to update the image pixel values. The process continues either for a prespecified number of iterations or until the magnitude of the changes from one iteration to the next falls below some prespecified threshold. One popular form of the iterative update procedure is given by the following equation

$$\vec{f}^{(k+1)} = \vec{f}^{(k)} + \frac{g_{i_k} - <\vec{H}_{i_k}, \vec{f}^{(k)}>}{<\vec{H}_{i_k}, \vec{H}_{i_k}>} \vec{H}_{i_k}, \tag{1}$$

where $\vec{x}^{(k)}$ and $\vec{x}^{(k+1)}$ represent the current estimate of the image vector after the kth and $(k+1)$st iterations, respectively, $i_k = [k \bmod I+1]$, I is the total number of projection rays, g_{i_k} is the measurement for the i_kth projection ray, $\vec{H}_{i_k}$ is the transpose of the i_kth row of the projection matrix H relating the image vector to the measurement vector, and $<>$ denotes an inner product [1, chapter 11].

We implemented an ART algorithm that made use of equation (1). The i, jth element of the projection matrix H was chosen to equal the intersection length between the ith pixel and the jth projection ray. To calculate these intersection lengths, we used an efficient algorithm, due to Siddon [2], in which the intersections of rays with separate grids of parallel horizontal and vertical lines are computed, the intersections ordered appropriately, and the differences computed. In the interest of storage efficiency, we stored the weights corresponding only to a single projection view at any one time, recomputing them anew at each iteration. To further reduce storage requirements, we only stored non-zero weight elements along with a separate list of their corresponding indices.

We tested the algorithm by reconstructing images of a numerical breast phantom containing large, low-contrast background structures, as well as two small lesions, one centrally and the other peripherally located. Because all of the structures are ellipses, the phantom's projections could be computed analytically, and we generated sinograms of 64 angles × 64 bins and 16 angles × 64 bins. We reconstructed each of these using ART and FBP (using a ramp filter with cutoff at the Nyquist frequency). We then added Poisson noise (200,000 total counts in the 64-angle sinogram, 50,000 counts in the 16-angle sinogram) and reconstructed using the same techniques.

2.1.2 Results and conclusions—ART

The results of the ART and FBP reconstructions are illustrated in Fig. 1. The ART reconstructions are obviously disappointing and especially so in the presence of noise. Both methods perform well for the noise-free reconstructions from 64 angles, but in the presence of noise, it is more difficult to discern the lesions in the ART reconstruction than in the FBP image. Moreover, the ART reconstruction contains noise structures that could be mistaken for lesions. In the noise-free, 16-angle reconstructions, the lesions are again more visible in the FBP reconstruction despite the presence
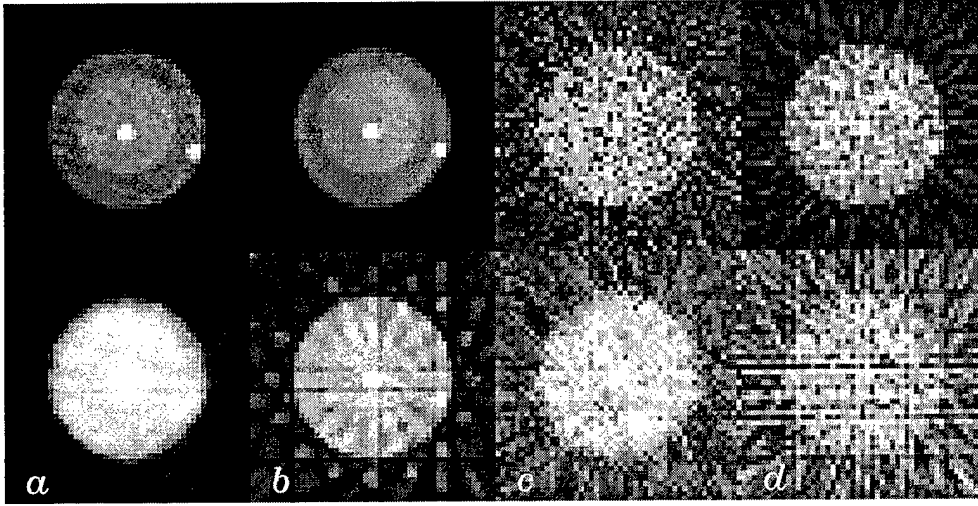
Figure 1: Reconstructions of a numerical breast phantom using ART and FBP. The top row corresponds to a 64-angle sinogram and the bottom row to a 16-angle sinogram. Column (a) is reconstructed by ART from noise-free data, column (b) by FBP from noise-free data, column (c) by ART from noisy data, and column (d) by FBP from noisy data.

of distracting star artifacts. Finally, while the lesions are difficult to discern in either of the noisy 16-angle reconstructions, the FBP reconstruction again seems superior. Coupled with the fact that ART is much more computationally intensive than FBP, these results are unencouraging.

The reasons for ART's poor performance in the noise-free case may lie in the combination of numerical instability and the lack of regularization in the straightforward ART implementation of equation (1). Clearly these factors are only exacerbated in the presence of noise. Attempts to improve the quality of the reconstructions by incorporating explicit statistical information as well as prior assumptions about the simplicity, symmetry, and near-binary nature of the expected reconstructions met with numerous difficulties. The most natural constraints did not lend themselves to efficiently implementable mathematical formulations and approximations of the constraints led to computationally intensive algorithms with little apparent advantage over FBP. These poor results prompted us to modify our approach to tasks 2 and 3, leading us to develop and characterize algorithms that are alternatives to ART rather than extensions of it.

2.2 Task 2—Characterizing angular sampling requirements

Using numerical simulation, systematically explore the dependence of reconstruction image quality on the number of projections and the complexity of the input image.

In this task, we aimed to characterize angular sampling requirements for dedicated SPECT SMM. The minimum number of angular views required to produce an accurate tomographic reconstruction of a given object is dictated by two factors. First, the angular sampling of the object's sinogram must satisfy, at least approximately, the Nyquist sampling condition. Absent this, any reconstruction is doomed to suffer from angular aliasing artifacts. Second, the number of angular samples must satisfy any implicit assumptions made by the reconstruction algorithm about the density of angular sampling. The surprising success of FBP in the comparisons with ART in task 1, even for few-view reconstructions, prompted us to reconsider its application to the few-view reconstruction problem and, in the context of this task, to investigate its particular angular sampling requirements. That the FBP and Nyquist sampling requirements are not necessarily equivalent can be appreciated most keenly when considering the case of imaging a

circularly symmetric object. In this instance, a single projection view is sufficient to satisfy the Nyquist sampling condition, while a FBP reconstruction from this single view would be an uninterpretable set of parallel streaks. The significance of this investigation is that if the Nyquist condition were satisfied by a number of samples less than the number required by FBP, it would be in principle possible to interpolate the additional views needed to eliminate the star artifacts without compromising the accuracy of the images or introducing new artifacts.

2.2.1 Methods—Angular sampling requirements in FBP

In [3], Brooks *et al.* present an elegant analysis of FBP's angular sampling requirements for reconstructing a circularly symmetric object without significant angular-sampling artifacts. This requirement provides a lower bound on the number of angular views needed to reconstruct a more general, asymmetric object without significant angular-sampling artifacts. The analysis also provides insight into the nature of the artifacts that arise when these requirements are not met. They show that the error term, or the difference between the exact reconstruction and the reconstruction from M views, is given by

$$\epsilon(r,\theta) \cong 2\pi(-1)^M \cos(2M\theta) \int_{-\infty}^{\infty} \widetilde{P}(\nu) J_{2M}(2\pi\nu r)d\nu, \tag{2}$$

where r, θ are polar coordinates in image space, $\widetilde{P}(\nu)$ is the Fourier transform of the filtered projections (they are all, of course, identical for a circularly symmetric object), and $J_{2M}(x)$ is a Bessel function of the first kind of order $2M$. A few features of this error term bear mention. First, the $\cos(2M\theta)$ factor, which represents an angular modulation, accounts for the star-like appearance of few-view FBP reconstructions. Second, the radial oscillation of the Bessel function accounts for the broken appearance of the streaks comprising the star. Finally, the Bessel function $J_{2M}(x)$ is essentially zero until its argument becomes approximately equal to its order. Thus, Brooks *et al.* argue, the error term is negligible if the order $2M$ exceeds the maximum value of the argument by about 10%. The maximum value of this argument obtains when $\nu = \nu_{\max} = \frac{L}{4R_\xi}$ and $r = R_\xi$, where L is the number of projections bins and R_ξ the radius of the field of view, and thus they find that a reasonable estimate of the number of views needed over π to effectively eliminate angular-sampling artifacts is given by

$$M \geq \frac{1.1\pi L}{4}. \tag{3}$$

Again, because this result was obtained for a circularly symmetric object, it should be viewed as a lower bound for the number of views needed for reconstruction of an arbitrary, asymmetric object.

2.2.2 Methods—Nyquist angular sampling requirements for breast phantoms

Given this result for FBP's angular sampling requirements, we turned to an analysis of the sampling requirements on the angular component of the sinogram of a model breast phantom. We made use of the circular sampling theorem (CST), which states that a function $f(\theta)$ that is periodic with period 2π and bandlimited to frequency $\nu_c = K/2\pi$ (i.e., the coefficients F_k of the function's Fourier series satisfy $F_k = 0$ for $|k| > K$) can be recovered exactly from its samples $f(2\pi j/MN)$, $j = 0, \ldots, N-1$, if $N > 2K$ [4, 5].

We modeled the breast phantom as consisting of an outer boundary represented by a circle of radius R centered at the origin and a lesion, modeled as a Gaussian profile of radial width (standard deviation) R_0, centered at a position $(x_0, 0)$. Thus the activity distribution $t(x, y)$ can be described by

$$t(x,y) = A_b \text{circ}\left(\frac{\sqrt{x^2+y^2}}{R}\right) + \frac{A_l}{2\pi R_0^2} \exp\left\{-\frac{[(x-x_0)^2 + y^2]}{2R_0^2}\right\}, \tag{4}$$

where $\text{circ}(r/R)$ denotes a circle of radius R, and A_b and A_l are the background and lesion activity concentrations, respectively. Placing the lesion on the x-axis simplified the analysis but did not affect any of the conclusions as they

7

must necessarily be invariant under a rotation of the phantom. The off-center lesion determined the angular sampling requirements, as the outer circle, being circularly symmetric, produces a sinogram with no angular variation at all.

The most straightforward way to approach the problem would have been to compute analytically the sinogram of this off-center lesion and then to have expanded the angular component in a Fourier series to determine its angular bandwidth. Unfortunately, this tack is mathematically intractable. Fortunately, by invoking the central slice theorem, an equivalent analysis could be performed by computing the 2D Fourier Transform (FT) of the off-center lesion, expressing the result in polar coordinates, and computing the Fourier series of the azimuthal variable.

The basic outline of this approach is given here. The 2D FT of an unshifted Gaussian (i.e., with $x_0 = 0$) is given simply by

$$T_{x_0=0}(\nu_x, \nu_y) = A_l R_0^2 \exp\left\{-2\pi^2 R_0^2(\nu_x^2 + \nu_y^2)\right\}, \tag{5}$$

where ν_x and ν_y are the frequencies conjugate to x and y, respectively. By the Fourier shift theorem, the 2D FT of an off-center lesion differs from that of a centered lesion only by a phase factor:

$$T(\nu_x, \nu_y) = A_l R_0^2 \exp\left\{-2\pi^2 R_0^2(\nu_x^2 + \nu_y^2)\right\} \exp(-j2\pi\nu_x x_0). \tag{6}$$

Expressing this in polar coordinates (ν, θ) yields

$$T(\nu, \theta) = A_l R_0^2 \exp\left(-2\pi^2 R_0^2 \nu\right) \exp(-j2\pi x_0 \nu \cos\theta). \tag{7}$$

Now we wish to compute the Fourier series coefficients for an expansion in θ:

$$\begin{aligned}
T_k(\nu) &= \frac{1}{2\pi} \int_0^{2\pi} T(\nu, \theta) e^{-jk\theta} d\theta \tag{8} \\
&= A_l R_0^2 \exp\left(-2\pi^2 R_0^2 \nu\right) \left[\frac{1}{2\pi} \int_0^{2\pi} e^{-j2\pi x_0 \nu \cos\theta - jk\theta} \right] d\theta. \tag{9}
\end{aligned}$$

The factor in square brackets is given by $(-j)^k J_k(2\pi\nu x_0)$, so

$$T_k(\nu) = A_l R_0^2 \exp\left(-2\pi^2 R_0^2 \nu\right) (-j)^k J_k(2\pi\nu x_0). \tag{10}$$

It is well known that $J_k(x)$ is negligible for $|k| > |x|$, and thus we conclude that $T_k(\nu)$ is negligible for $|k| > 2\pi\nu x_0$. Thus by the circular sampling theorem, we require $N > 4\pi\nu_{\max}x_0$ samples, or $M > 2\pi\nu_{\max}x$ projections (each projection provides two azimuthal samples in frequency-space) to adequately sample the angular portion of this object's sinogram for all relevant ν. The highest significant frequency $\nu_{\max}$ is determined by identifying the ν beyond which $T_k(\nu)$ falls below some cutoff. This is determined principally by the $\exp\left(-2\pi^2 R_0^2 \nu\right)$ factor in Eq. 10, which is a Gaussian of standard deviation $1/(2\pi R_0)$. If we set the cutoff $\nu_{\max}$ to be the point at which the Gaussian falls to 1% of its peak value,

$$\nu_{\max} \cong \frac{3.0}{2\pi R_0}, \tag{11}$$

we thus conclude that

$$M > 3.0 \frac{x_0}{R_0} \tag{12}$$

is the number of projections required over π for reasonable satisfaction of the Nyquist condition on the angular component of the sinogram.

To get a sense of the requirement imposed by this expression, consider the SPECT phantom experiments to be discussed in Sec. 2.8.1, with a physical phantom comprising a cylinder of radius 7 cm containing a spherical lesion of radius 0.5 cm placed about 5 cm from the axis of the cylinder, which is itself aligned with the SPECT system's axis of rotation. While the true 2D cross section of a spherical lesion is obviously a circle, the finite radial resolution of the system means that the Gaussian model is likely more appropriate for the "effective" lesion being reconstructed, and,

moreover, that its effective radius R_0 should be estimated by adding its actual radius in quadrature with the standard deviation of the system's resolution function, assumed to be a Gaussian. For a system resolution of FWHM 1.5 cm, we find $R_0 \approx \sqrt{0.5^2 + (1.5/2.35)^2} = 0.81$ cm, where we have used the fact that for a Gaussian the FWHM$\approx 2.35\sigma$. For $x_0 = 5$ cm, this means that we require $M > 19$ projections to satisfy reasonably the angular sampling requirement of Eq. 12. By comparison, because $L = 128$ bins on this system, Eq. 3 predicts that FBP requires on the order of 110 projections simply to reconstruct the outer circle accurately; its requirements for accurate reconstruction of the lesion may be higher still. Thus, in this case, the Nyquist sampling condition on the angular part of the sinogram can be satisfied by a smaller number of angles than is needed to produce FBP images that are essentially free of angular sampling artifacts.

A few caveats should accompany this argument. In the experiment discussed in Sec. 2.8.1, a full-size SPECT system is being used to image a relatively compact phantom. In practice, the field of view and the number of detector elements could be halved and still encompass the object, but even in this case, with $L = 64$, the FBP requirement is still on the order of 55 projections and thus still higher than the Nyquist requirement. A second potential counterargument is that the analysis of Eq. 2 indicates that the error for imaging a circularly symmetric object tends to get shifted toward the periphery of the field of view as the number of views are increased, so it could be argued that the FBP error introduced by the large outer circle would not necessarily impede detectability of the lesion. However, this neglects the fact that the off-center lesion itself will likely radiate streak artifacts in a few-view FBP reconstruction. These may not overlap itself, but they could create the appearance of additional lesions elsewhere in the phantom, or, if there is genuinely a second lesion present, interfere with its detection.

In any case, the argument developed in this section at least suggests that there are times, especially when imaging relatively simple objects, when the Nyquist sampling condition for the angular component of the sinogram can be satisfied by a smaller number of angles than is needed to produce FBP images that are essentially free of angular sampling artifacts. In these situations, the additional projections needed by FBP can, in principle, be interpolated from the measured ones.

2.3 Task 3—Modifying the ART approach

Modify the algorithm to perform a purely binary reconstruction of the MIBI distribution.

As discussed above, the poor performance of the ART algorithm discovered with the completion of task 1 lead us to develop algorithms that are alternatives to ART rather than extensions of it. The analysis of angular sampling requirements carried out in task 2 suggested one approach—interpolating additional views in the few-view sinogram prior to reconstruction by FBP. Because the Nyquist sampling requirements for a typical breast image are modest, it may be possible to eliminate FBP's star artifacts by use of such interpolation without compromising the accuracy of the images or introducing new artifacts. In this section we describe the various interpolation approaches considered as well as our efforts to characterize their accuracy and noise properties.

2.3.1 Methods—Sinogram interpolation

The interpolation method used should of course exploit the periodicity of the data in the angular coordinate, and we have explored four popular periodic interpolation approaches: linear interpolation with periodic boundary conditions, periodic spline interpolation, zero-padding (ZP) interpolation, and circular-sampling theorem interpolation (CST). Linear interpolation needs no introduction, and ZP is also quite familiar: it involves extending the discrete Fourier transform (DFT) of a finite sequence with zeroes and then taking an inverse DFT to obtain a more densely sampled version of the sequence, with values interpolated at intermediate positions between the original measured samples [6, 7, 8]. CST interpolation is a special case of Whittaker-Shannon (W-S) sinc interpolation that applies to periodic functions. The CST holds that if a function $g(x)$ is periodic with period X, it can be reconstructed from a finite number N of samples taken over one period. Specifically, consider $g(x)$ to be sampled such that $g_n = g(n\Delta x)$, where $\Delta x = X/N$ and $n = 0, \ldots, N - 1$. If $g(x)$ is bandlimited to frequency K (i.e., the coefficients of expansion a_k of the

function's Fourier series satisfy $a_k = 0$ for $|k| > K$), and if $N \geq 2K + 1$ and is odd, then $g(x)$ can be reconstructed exactly from its samples by use of

$$g(x) = \sum_{n=0}^{N-1} g_n \frac{\sin\left[\frac{\pi}{\Delta x}(x - n\Delta x)\right]}{N \sin\left[\frac{\pi}{X}(x - n\Delta x)\right]}. \tag{13}$$

Similarly, if $N \geq 2K$ and is even, then the value of $g(x)$ may be determined exactly at arbitrary x using

$$g(x) = \sum_{n=0}^{N-1} g_n \left\{ \frac{\sin\left[\frac{\pi}{X}(N-1)(x - n\Delta x)\right]}{N \sin\left[\frac{\pi}{X}(x - n\Delta x)\right]} + \frac{1}{N} \cos\left[\frac{\pi}{\Delta x}(x - n\Delta x)\right] \right\}. \tag{14}$$

As for periodic spline interpolation, it involves fitting the data with piecewise cubic polynomials that are continuous up to and including the second derivative at the joints between pieces [9, 10]. A spline $g(x)$ can be represented by

$$g(x) = a_n + b_n(x - x_n) + c_n(x - x_n)^2 + d_n(x - x_n)^3, \quad x \in [x_n, x_{n+1}], \tag{15}$$

and it can be shown that the coefficients a_n, b_n, c_n, and d_n can be obtained from the vector $\mathbf{g}$ of measured data points through matrix multiplications of the form $\mathbf{a} = \mathbf{A}\mathbf{g}$, where the matrices such as $\mathbf{A}$ are independent of the data and are constructed in accordance with periodic boundary conditions.

2.3.2 Results—Connections among interpolation methods

The first aspect of the interpolation work was to explore the little-understood connections among ZP, CST, and W-S interpolation. We showed that they are in fact mathematically equivalent for the task of increasing the density of angular samples, with ZP interpolation being by far the most computationally efficient. The proof of these equivalences is given in the paper entitled, "Mathematical equivalence of zero-padding interpolation and circular sampling theorem interpolation with implications for direct Fourier image reconstruction," which is included as an appendix to this report.

Interestingly, while we are primarily concerned with the interpolation of additional projection views in the spatial domain for use in FBP reconstruction, that paper also discusses the application of the interpolation approaches to an alternative reconstruction approach: direct Fourier reconstruction. Direct Fourier techniques are based on the central slice theorem, which states that the Fourier transform of each projection view corresponds to a line through the origin in the Fourier transform space of the object being imaged. The Fourier transforms of a set of projection views then provides a set of polar samples of the object's Fourier transform. If the samples are interpolated onto a Cartesian grid, the object can then be reconstructed by use of the 2D fast Fourier transform. Researchers had previously proposed using the CST to perform the azimuthal interpolation needed to make that coordinate transformation [11, 12], but our work established that an alternative approached, proposed by Pan and Kak [13], in which ZP interpolation was used to increase the density of samples in the azimuthal direction prior to the use of linear interpolation to make the final jump onto the Cartesian grid, was nearly equivalent and much more efficient. Again, details are given in the paper "Mathematical equivalence of zero-padding interpolation and circular sampling theorem interpolation with implications for direct Fourier image reconstruction," which is included as an appendix to this report.

2.3.3 Results—Comparative accuracy of interpolation methods

Having established the equivalence of ZP and CST interpolation, we then dropped the CST from further consideration (owing to ZP's computational advantages) and turned to the task of comparing the accuracy of the remain approaches. Specifically, we have compared the accuracy of the ZP and periodic spline interpolation approaches, as well as linear interpolation with periodic boundary conditions, for the task of interest: interpolating additional projections in a few-view sinogram [14]. Simply comparing the success of the approaches in interpolating a single sinogram each for one or two canonical phantoms would have provided more anecdotal than genuinely rigorous evidence on which

to base the choice of interpolation method for few-view tomography. The outcome could have depended as much on numerical happenstance as on the genuine strengths of the approaches. Instead, we generated 100 different "realizations" of each of the two types of numerical phantom—Shepp-Logan and breast—by choosing the parameters specifying the constituent ellipses of each type to vary according to predetermined probability laws. Corresponding sinograms of 128 bins×1024 projection views were computed analytically and subsampled to 16, 32, 64, 128, 256, and 512 projection views. Each subsampled sinogram was interpolated to 1024 projection views by each of the methods under consideration and the normalized root-mean-square-error (NRMSE) with respect to the true 1024 projection view sinogram computed. In addition, images were reconstructed from the interpolated sinograms by FBP and the NRMSE with respect to the true phantom computed. The non-parametric signed rank test was then used to assess the statistical significance of the pairwise differences in mean NRMSE among the interpolation methods for the various conditions: phantom family (Shepp-Logan or breast), number of measured projection views (16, 32, 64, 128, 256, or 512), and endpoint (sinogram or image). Periodic spline interpolation was found to be superior to the others in a statistically significant way for virtually every condition. Further details are provided in the paper "Comparison of angular interpolation approaches in few-view tomography using statistical hypothesis testing," which is included as an appendix to this report. Tables 1-4 and Figs. 2 and 3 of that paper contain the principle results summarized here.

2.3.4 Results—Comparative noise properties of interpolation methods

We have also calculated analytic expressions for the propagation of noise from the measured samples into the interpolated curves for the methods under consideration. Consider a periodic function $g(x)$ having period X that is sampled N times over one period, i.e. at points $x_n = Xn/N$ for $n = 0, \ldots, N - 1$. Assume each measurement is corrupted by additive, zero-mean noise. The measured samples of $g(x)$ can then be represented as

$$\mathbf{g}(x_n) = \langle \mathbf{g}(x_n) \rangle + \mathbf{n}(x_n), \tag{16}$$

where bold-faced type denotes a random variable, $\langle \rangle$ represents the expectation operator and $\mathbf{g}(x_n)$ is the additive, zero-mean noise. Let $\widehat{\mathbf{g}}(x)$ be the curve interpolated from the noisy samples of equation (16). The covariance of this function for two points x and x' is defined by

$$\mathrm{cov}(x, x') = \langle [\widehat{\mathbf{g}}(x) - \langle \widehat{\mathbf{g}}(x) \rangle] [\widehat{\mathbf{g}}(x') - \langle \widehat{\mathbf{g}}(x') \rangle] \rangle. \tag{17}$$

Of course, once the covariance has been computed, the variance at any point x is given by $\mathrm{var}(x) = \mathrm{cov}(x, x)$.

Evaluation of equation (17) was straightforward for CST and ZP interpolation, but more subtle for spline interpolation. Details of the evaluations are given in the paper entitled "Noise properties of periodic interpolation methods with implications for few-view tomography," which is included as an appendix to this report. Fourier-based interpolation has the favorable noise property that when it interpolates among samples contaminated with zero-mean white noise of a given variance, the resulting interpolated curve has stationary noise with the same variance. When interpolating among samples contaminated with Poisson noise, the resulting interpolated curve has variance that is flat locally and that tracks the noise in the measured samples over longer distances. These properties are desirable because wide fluctuations in noise levels in the interpolated curve can lead to artifacts in reconstructed images. Linear interpolation does lead to such wide fluctuations, with the variance at points midway between measured samples falling to half the level of the variance in the measured samples when interpolating among samples contaminated with zero-mean white noise. The noise properties of spline interpolation lie somewhere between those of Fourier-based and linear interpolation. Specifically, when interpolating among samples contaminated with zero-mean white noise, the resulting interpolated curves display only a slight dip in variance between the positions of the measured samples. Similar local behavior is observed in the case when the samples are contaminated by Poisson noise.

The results just discussed of course apply directly to the characterization of noise properties of sinograms in which additional views have been interpolated by the methods under consideration. However, it is just as important to consider the noise properties of the *images* reconstructed from such sinograms, because these noise properties, as characterized by a noise power spectrum (NPS), greatly influence the detectability of small objects [15]. For this

reason, we wished to calculate the NPS of images reconstructed after sinogram interpolation by ZP and periodic spline interpolation, as well as the NPS of images reconstructed directly from both a small and large number of views.

Although the noise power spectrum is strictly defined only for stationary noise processes, and the noise in images reconstructed from interpolated sinograms is not stationary due to the correlation between projections, there is precedent for examining the so-called average power spectrum of nonstationary processes [16]. Details of the methods are given in the paper "Noise properties of periodic interpolation methods with implications for few-view tomography," which appears as an appendix to this report, and the results are given in Fig. 5 of that paper, where it is seen that the ZP NPS is much more similar to the NPS of images reconstructed from a full complement of 120 original views than was the spline NPS. Indeed, the spline NPS demonstrated the existence of non-uniform angular correlations that could potentially hinder detection tasks.

2.3.5 Conclusions—Sinogram interpolation

We have thus conducted extensive investigations of the connections among, as well the accuracy and noise properties of the methods under consideration for interpolation of additional views in a few-view sinogram. It is perhaps worth summarizing the important findings and conclusions.

- We have shown that ZP interpolation and CST interpolation are mathematically equivalent insofar as the discrete sets of values yielded by ZP interpolation are the same as would be yielded at those points by CST interpolation.

- We have compared the accuracy of ZP, periodic spline, and linear interpolation satisfying periodic boundary conditions for the interpolation of additional views in a few-view sinogram and found that periodic spline interpolation is almost always the most accurate in practice. Despite its theoretical exactness, zero-padding tends to produce ringing artifacts when faced with mild violations of its assumptions about the data.

- We have studied the noise properties of curves interpolated by means of the three methods as well as images reconstructed from sinograms with additional views interpolated by means of the methods. ZP interpolation has, in general, the best performance because it yields curves and images with more stationary noise than those produced by the other approaches.

- This noise finding should probably carry less weight than the accuracy finding, because, in general, it is best to mitigate noise by smoothing each measured projection *prior* to the interpolation of additional views. Thus, on the basis of the accuracy studies, periodic spline interpolation is our preferred method for the generation of additional views in a few-view sinogram. The question of how best to smooth the measured projections is the topic of the next task, which involves the development of a Bayesian approach to that problem.

Technical Objective 2

To develop Bayesian algorithms that incorporate prior knowledge about the distribution being imaged, in particular the knowledge about the boundary of the suspected lesion extracted from the binary lesion of objective 1.

2.4 Task 4—Develop a Bayesian algorithm

Develop a Bayesian algorithm that incorporates prior knowledge about the MIBI distribution.

Bayesian algorithms refer to algorithms that maximize some sort of posterior distribution comprising two terms: a likelihood term reflecting the distribution of the measured data conditional on the underlying quantity of interest and a regularizing term reflecting some prior beliefs about the likely distribution of the underlying quantity of interest itself. While Bayesian algorithms have often figured in medical tomography as a means of regularizing the fully iterative maximum-likelihood expectation-maximization (MLEM) algorithm, we have developed an alternate form of

12

Bayesian algorithm that operates purely in the projection domain and incorporates information about the expected smoothness of the projections. Reconstruction then proceeds by use of FBP. The approach leads to statistically significant improvements in resolution/variance tradeoffs relative to those achievable by use of the projection-smoothing approaches usually used in concert with FBP. Moreover, a generalization of the approach allows for a degree of control over resolution uniformity and isotropy properties that is quite difficult to achieve in the context of fully iterative Bayesian approaches.

2.4.1 Methods and results—A Bayesian, roughness-penalized nonparametric regression projection smoothing approach

We have developed a projection-smoothing approach that is a form of Bayesian, roughness-penalized nonparametric regression. It entails fitting to the measured projection samples at each view smooth curves that maximize an objective function that comprises two terms. The first term is a Poisson likelihood and the second is a term that penalizes excessively jagged fit curves. Thus the prior information about the object that is being incorporated are that its projections are relatively smooth. The smoothed curves are then resampled and reconstruction proceeds by use of FBP without any additional filtration (other than the requisite ramp). This Bayesian smoothing step is thus an alternative to the usual apodization of the ramp filter used to mitigate noise in FBP. The shortcoming of apodization is that it is equivalent to convolution with a shift-invariant filter in the projection domain. This is optimal only for data contaminated by stationary noise, whereas noise in emission tomography follows a Poisson distribution and is, in general, nonstationary. The explicit modeling of the Poisson statistics in our approach leads to statistically significant improvements in the noise/resolution tradeoffs relative to that achievable by use of simple shift-invariant apodization filters.

The basic technique is introduced in the paper entitled "Few-view tomography using roughness-penalized nonparametric regression and periodic spline interpolation," which is included as an appendix to this report. Each set of projection samples $g_i = p^{(\phi_j)}(\xi_i)$, $i = 0, ..., N-1$, where ϕ_j denotes the projection view, ξ_i the projection bin, and N the number of bins per projection, is fit with a curve $\widehat{p}^{(\phi_j)}(\xi_i)$ maximizing a roughness-penalized Poisson likelihood function

$$\Phi(p, g) = \sum_{i=0}^{N-1} [g_i \ln p(\xi_i) - p(\xi_i)] - \alpha \int [p''(\xi)]^2 \, d\xi, \tag{18}$$

where $''$ denotes the second derivative, α is the smoothing parameter, and where, for simplicity, we have dropped the dependence on ϕ_j. It can be shown that this objective function is always maximized by a natural cubic spline, which is simply a cubic spline constrained to be linear beyond the first and last measurement points. The coefficients of the spline can be found through a computationally efficient iterative procedure.

A very interesting generalization of the technique, along with substantial characterization of its resolution properties, is described in the paper entitled "Nonparametric regression sinogram smoothing using a roughness-penalized Poisson likelihood objective function," which appears as an appendix to this report. The generalization involves enforcing the smoothness constraint not on the underlying function itself, but rather on some monotonic transformation of it. This generalization makes it possible to exert control over resolution uniformity and isotropy properties, and specifically, through the choice of a square root link function, to achieve essentially uniform and isotropic resolution in reconstructed images with little sacrifice is resolution-variance tradeoffs. This is a very desirable outcome that is difficult to achieve in the context of fully iterative Bayesian approaches. Standard quadratic penalties in the fully iterative context lead to decidedly nonuniform resolution and the design of penalties that lead to uniform and isotropic resolution is complicated and time consuming. All of these issues are discussed in greater detail in the paper, "Nonparametric regression sinogram smoothing using a roughness-penalized Poisson likelihood objective function," but in short, the proposed approach has some clear advantages that can be summarized as follows: It has statistically significant resolution/noise advantages over apodization windows, it captures some of the statistical benefits of a fully iterative reconstruction algorithm, at a fraction of the computational cost, and it also can also be used to produce images having uniform and isotropic resolution more naturally than when making use of fully iterative algorithms.

2.4.2 Methods—Direct spline reconstruction

When smoothing splines are fit to each projection in the nonparametric regression technique, they must be resampled to yield a discrete sinogram if reconstruction is to proceed by FBP. Discarding the continuous information embodied in the spline coefficients seems wasteful because after filtration of these samples the FBP algorithm interpolates among them during backprojection. We have investigated an alternative to FBP in which the reconstructed image is expressed explicitly in terms of the coefficients of the splines fit to the projections. While evaluating this expression is computationally intensive, the approach is found to yield higher resolution images than does FBP. This approach is described in detail in the paper "Spline-based inverse Radon transform in two and three dimensions," which is included as an appendix to this report. Unfortunately, the intensive computational demands of the resulting algorithm for 2D reconstruction (it fares much better in 3D) preclude its incorporation into the family of algorithms under development.

Technical Objective 3

To use numerical simulations to determine the optimal number of projection angles to be used by each of the algorithms developed, from the standpoint of maximizing diagnostic accuracy while minimizing image acquisition time.

2.5 Task 5—Characterize diagnostic accuracy of few-view reconstructions

Define a cost function that represents the tradeoff between improving diagnostic accuracy and reducing image acquisition time. Apply the function to all the algorithms developed in tasks 3 and 4.

The aim of this task was to establish whether the algorithm developed—the combination of Bayesian projection smoothing and spline-based interpolation of additional projections—manages to preserve diagnostic information that is present in images reconstructed from standard numbers of views when reconstructing from smaller numbers of views. To this end, we chose to image a cardiac phantom containing a 1-cm defect insert and placed at increasing radial offsets from the center of rotation of a SPECT system. Placing the phantom at increasing radial offsets increases the bandwidth of the angular functions at each projection bin and thus makes the interpolation task more challenging. The diagnostic information of interest in this situation is the detectability of the defect, and as a figure of merit we generated bullseye plots [17] from each set of reconstructions. These are remappings of the reconstructed cardiac phantom images that make it very clear whether or not a defect is detectable in the reconstructed data or not.

2.5.1 Methods—Assessment of diagnostic accuracy

To carry out this task, projections were acquired of a Data Spectrum ventricular phantom placed at five different radial offsets—0, 5, 9, 12, and 15 cm—from the center of rotation of a Picker 3000XP three-headed SPECT system fit with low-energy, high-resolution, parallel-hole collimators. The phantom was filled with 121 MBq (3.27 mCi) of Tc-99m, and contained a 1-cm defect insert. The acquired studies contained 120 projections over 360°. A 25-cm radius circular orbit and step-and-shoot mode was used for all of the acquisitions; each head acquired to a 128×128 pixel image, although only the 32 slices spanning the phantom were preserved. A total of about 500,000 counts was collected. From this data 3D sinograms were extracted corresponding to 15, 30, 60, and 120 projections, respectively. Thus we had 20 different sinograms, corresponding to the 20 possible combinations of radial offset and number of angular views. We reconstructed images from these 20 sinograms using four different processing techniques:

1. No pre-smoothing of the sinogram and slice-by-slice reconstruction from available projections by FBP using a Hanning filter (cutoff=0.8).

2. No pre-smoothing of the sinogram, spline interpolation from the available projections to 120 projections, and slice-by-slice reconstruction by FBP using a Hanning filter (cutoff=0.8).

14

3. Spline-based, roughness-penalized Poisson likelihood smoothing of the sinogram (using the identity link) and slice-by-slice reconstruction from the available projections by FBP using a ramp filter (cutoff=1.0).

4. Spline-based, roughness-penalized Poisson likelihood smoothing of the sinogram (using the identity link), periodic spline interpolation from the available projections to 120 projections, and slice-by-slice reconstruction by FBP using a ramp filter (cutoff=1.0).

In applying the roughness-penalized nonparametric regression smoothing, the smoothing parameter λ was selected by an automatic procedure known as cross-validation, modified to account for the Poisson statistics [18, 19, 20, 21].

2.5.2 Results—Assessment of diagnostic accuracy

For ease of comparison and simplicity of presentation, the reconstructed images are grouped in Fig. 2 by the number of projections in the original sinogram, and are shown for only three of the radial offsets: 0, 9, and 15 cm. These three suffice to illustrate the trends observed. For each combination of number of projections and radial offset we show the results of reconstructing using each of the four techniques outlined above.

We observe that reconstructions from available projections without pre-smoothing or interpolation display star-shaped artifacts and a mottled appearance when the number of projections is small. Interpolation alone mitigates the star-shaped artifacts but leads to severe circular artifacts, particularly in the case of a small number of projections and a large radial offset. Smoothing alone reduces the noise visibility but has little effect on the star-shaped artifacts. The combination of smoothing and interpolation still produces circular interpolation artifacts in the case of a large radial offset combined with a small number of original views, but these are less severe than when interpolation alone was used. Overall, though, visually appealing reconstructions result for less challenging combinations of radial offset and number of projections, including as few as 15 projections in the 0 cm offset case.

While Fig. 2 suggests that the algorithms described can produce visually satisfactory reconstructions from relatively small numbers of projections, the most critical question is whether this can be achieved without hindering the detection of the small perfusion defects that is usually the goal of cardiac SPECT imaging. To answer this question we generated bullseye plots [17] from each set of reconstructions. If the reconstruction of the ventricle contained a uniform distribution of activity, the bullseye plot would be uniform. But if the activity was nonuniform, for instance if there was a perfusion defect having lower activity than surrounding areas, the appropriate sector of the dartboard would appear darker than the surrounding area. The bullseye plots corresponding to reconstructions from few-view sinograms that have been processed by the spline smoothing and interpolation techniques are shown in Fig. 3, along with bullseye plots corresponding to 120-angle sinograms reconstructed without spline processing. Our phantom contained a 1-cm perfusion defect insert, which produces a depression that is well resolved in many of the bullseye plots.

It is clear from these bullseye plots that the defect remains detectable for as few as 15 angles in the case of the 0 cm offset, as few as 30 angles in the case of the 9 cm offset, and as few as 60 angles in the case of the 15 cm offset. These findings correlate well with the visual appearances of the images in Fig. 2. The plots in which the defect is not visible correspond to the images in which severe interpolation artifacts are evident. These results indicate that the combined spline smoothing and interpolation approach has the potential to generate diagnostically useful images of simple objects from as few as 15 projections.

Technical Objective 4

To test the performance of the algorithms in reconstructing the distribution of MIBI in a breast phantom, using a miniature gamma camera already developed at the University of Chicago to acquire the projection images.

Owing to major personnel changes and redirection of research commitments, the development of nuclear medicine instrumentation at the University of Chicago, and along with support for the improvement and operation of the miniature gamma camera, has essentially ceased over the last two years. For this reason, it was not possible either to

15

	0 cm	9 cm	15 cm
No Processing			
Interpolation Alone			
Smoothing Alone			
Smoothing + Interpolation			

15 Angles

	0 cm	9 cm	15 cm
No Processing			
Interpolation Alone			
Smoothing Alone			
Smoothing + Interpolation			

30 Angles

	0 cm	9 cm	15 cm
No Processing			
Interpolation Alone			
Smoothing Alone			
Smoothing + Interpolation			

60 Angles

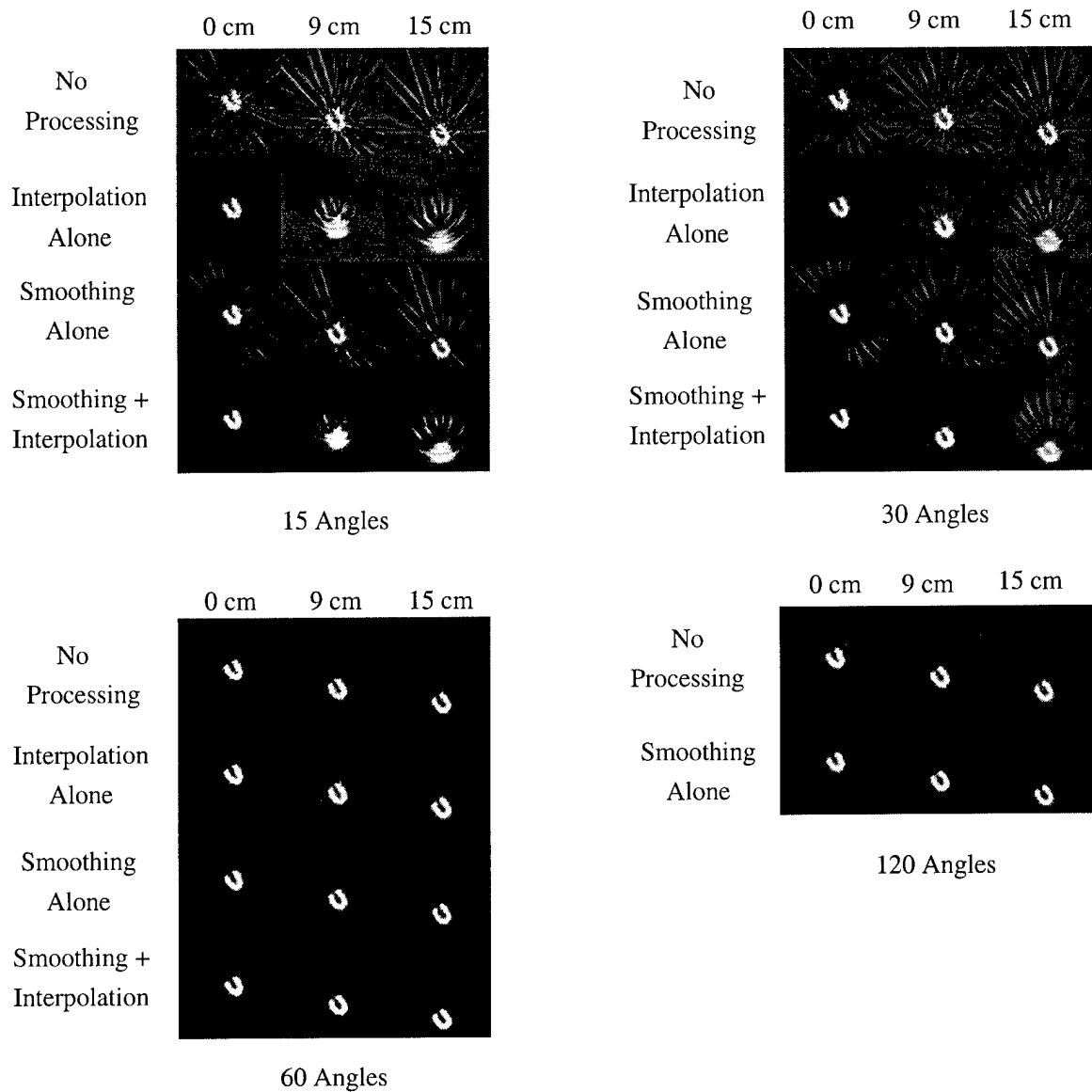| | 0 cm | 9 cm | 15 cm |
| No Processing | | | |
| Smoothing Alone | | | |

120 Angles

Figure 2: Reconstructions of a representative slice of a cardiac phantom image at three different radial offsets using four different preprocessing approaches and 15, 30, 60, and 120 projections.
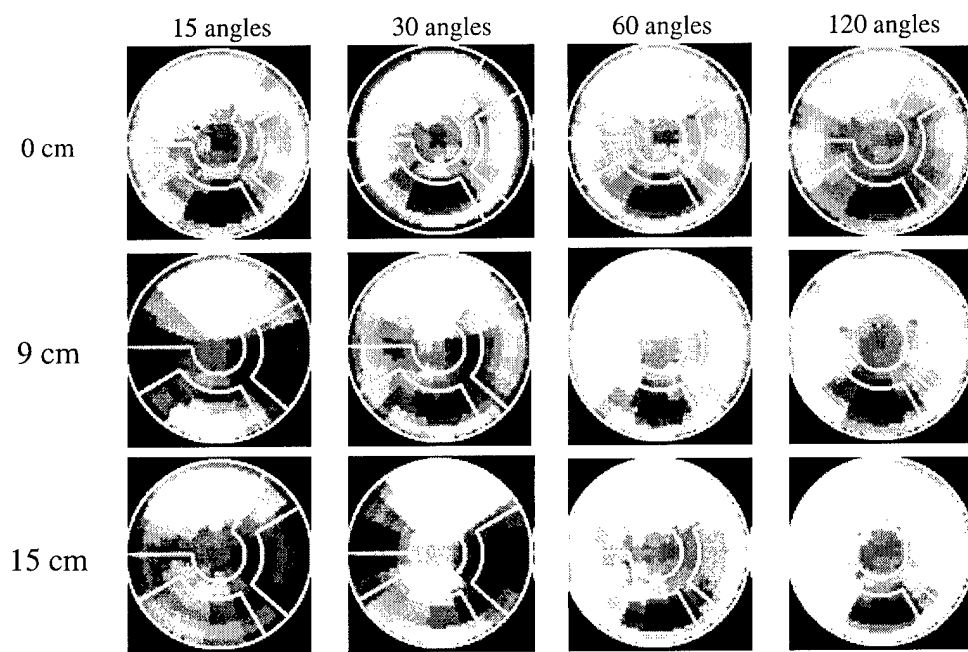
16

Figure 3: Bullseye plots for the cardiac phantom placed at various radial offsets and for four different numbers of projection angles. The reconstructions from 15, 30, and 60 angles used the nonparametric regression smoothing technique followed by periodic spline interpolation to 120 views prior to reconstruction by ramp-filtered FBP. The reconstructions from 120 angles simply entailed Hanning-filtered FBP.

construct an anatomically realistic breast phantom or to perform imaging studies with the miniature gamma camera. Nonetheless, we modified an existing phantom to provide a reasonable facsimile of the two-dimensional cross-section of a more realistic three-dimensional breast phantom, and we were able to simulate the dedicated SPECT SMM imaging geometry by operating a conventional three-headed SPECT system with the heads operating at the minimum possible radius of rotation. Thus, despite the setback, all of the tasks associated with this technical objective were pursued and achieved.

2.6 Task 6—Construction of a breast phantom

Construct a breast phantom capable of modeling the distribution of MIBI in normal and cancerous breasts as well as mimicking the attenuation properties of breast tissue.

For the reasons discussed above, the resources did not exist to construct an anatomically realistic breast phantom. Nonetheless, we assembled a reasonable facsimile of such by modifying a 14cm-diameter, 800 cc tub phantom to hold fillable, spherical lesion inserts of various sizes. Because the algorithms under consideration are all two-dimensional, we are primarily interested in having the projections of individual slices of the phantom be realistic. A two-dimensional slice through any section of this phantom containing the lesion comprises a large outer circle containing a certain concentration of activity and a smaller circular lesion containing a higher concentration of activity. This is precisely the cross-section that would have been simulated in a phantom having a more realistic, breast-like outer shell. The primary difference between the projections of lesion-containing slices in the cylindrical and breast-shaped phantoms would be the differences in scatter contribution from nearby regions of the phantom. However, because the algorithms developed are being evaluated primarily for their handling of sparsely sampled and noisy data, these differences in scattering do not fundamentally affect the conclusions of the study.

2.7 Task 7—Imaging the phantom using a simulated SPECT SMM system

Using the University of Chicago miniature camera, acquire projection images of the phantom to be used as input to the reconstruction algorithm.

The University of Chicago miniature camera project foundered before a camera of adequate field-of-view, stability, and resolution for dedicated SPECT SMM was developed. Nonetheless, we carried on with this aim by operating a standard SPECT system with the heads revolving at their minimum radius of rotation in order to simulate the potential resolution improvements possible in a dedicated SPECT SMM geometry (recall that resolution generally worsens with increasing distance from the collimator in a collimated system). Moreover, we expanded the scope of the task somewhat by seeking to verify the hypothesis that a dedicated breast SPECT geometry would indeed be better for SMM lesion detection than currently existing planar or conventional SPECT geometries. The experimental results of Wang *et al.* [22] suggested that it would be, but we undertook to answer the question more definitively and quantitatively by using the so-called ideal-observer framework to quantify the amount of information contained in the projections of a breast phantom obtained using the three different geometries.

2.7.1 Methods—Comparison of imaging geometries

The ideal-observer framework [15] offers a way of assessing the amount of information the data from an imaging device contain with regard to the performance of a specified task. For example, the simplest such task is the detection of a signal of known strength, shape, and location in a specified background. In this case, the framework seeks to quantify the degree to which an ideal observer—one who can use the information contained in the images to its fullest extent—can reliably distinguish images containing the background alone from images containing the background and the signal when both kinds of images are corrupted by noise, blurring, and other imperfections. For linear imaging processes in which the noise in the output image is assumed to be additive, Gaussian, zero-mean, stationary, and independent of the presence or absence of the signal, the ideal-observer framework allows us to characterize fully the

18

Planar SNR	Conventional SPECT SNR	Dedicated SPECT SNR
6.2	6.7	10.7

Table 1: Ideal-observer SNRs for detection of a 6:1, 1-cm lesion using planar, conventional SPECT, and dedicated SPECT scintimammography.

quality of the imaging system data with respect to the performance of the specified signal-detection task by a single number, the ideal-observer signal-to-noise ratio (SNR). Further details of its definition are given in the paper "Ideal-observer analysis of lesion detectability in planar, conventional SPECT, and dedicated SPECT scintimammography using effective multi-dimensional smoothing," which is included as an appendix to this report.

The images whose information content we wish to measure are not reconstructed images but rather the raw projection images acquired by the various imaging systems, for these offer the purest measure of the quality of the data acquired by the imaging system. The process of image reconstruction, while certainly helpful to human observers, can never increase the ideal-observer SNR and generally reduces it. Using reconstructed images to compare imaging geometries would thus conflate issues of data quality with issues of reconstruction algorithm accuracy. The methods we used to obtain the data necessary for computing ideal-observer SNRs in these circumstances are detailed in the attached paper "Ideal-observer analysis of lesion detectability in planar, conventional SPECT, and dedicated SPECT scintimammography using effective multi-dimensional smoothing,"

2.7.2 Results—Comparison of imaging geometries

The ideal-observer SNRs for the detection of a 1-cm lesion with a 6:1 lesion-background concentration ratio are listed in Table 1 for the three geometries. The SNR values given suggest that a dedicated SPECT geometry would lead to improved detectability for clinically typical lesions over the planar and conventional SPECT geometries, especially since in the presence of attenuation and scatter from the torso we would expect the difference between the dedicated and conventional SPECT geometries to be even greater than it is here. The success of the dedicated geometry can be attributed to the fact that it combines the advantages of the other two approaches: because of its small radius of rotation, it offers good sensitivity and resolution comparable to that of a planar view acquired with the scintillation camera flush against the phantom, while it offers the improved contrast offered by a tomographic system's ability to separate lesion activity from overlying and underlying activity.

2.8 Task 8—Evaluation of approaches applied to acquired data

Compare the reconstructed distributions to the known distributions of MIBI in the phantom.

Having established the superiority of the dedicated SPECT SMM imaging to other contenders, we turned to the application and evaluation of our developed algorithms to few-view dedicated SPECT SMM data. Two families of algorithms were considered: those based on a combination of Fourier-domain apodization for projection smoothing followed by ZP interpolation of additional projections, and those based on the Bayesian, spline-based projection smoothing followed by periodic spline interpolation of additional projections. The approaches were first used to reconstruct images from varying numbers of projections in order to assess the visual quality of the resulting images. The quantitative accuracy of the methods were then compared by computing resolution-variance curves for reconstructions of a numerical breast phantom.

2.8.1 Methods—Reconstructing images for few-view SPECT SMM

We imaged the cylindrical breast phantom described under task 6 containing a 1-cm outer diameter spherical lesion insert. This lesion size is representative of the smallest currently detected in SMM. We determined the concentration of Tc-99m to place in the background by assuming that 1% of a typical 25 mCi clinical dose of Tc-99m-sestamibi is

19

taken up by the myocardium. Using the volume of the myocardium in the Data Spectrum Corporation cardiac insert as a guide, along with the assumption that soft tissue will have a 1:15 concentration relative to the myocardium [22] allowed us to determine the expected concentration of activity in healthy breast tissue. We wished to simulate imaging for 30 minutes. Because this data was acquired in the context of another study requiring multiple realizations of the acquisitions, the calculated background concentration was scaled up by a factor of 30, and imaging time was reduced to one minute, thus maintaining clinically realistic count levels while accelerating the imaging process. Again, for reasons related to their use in another study, projection data were acquired first of the cylinder alone, without the lesion insert, and then of the lesion in a cold background. The lesion was filled with a relatively high concentration of activity and imaged for 30 minutes in order to acquire essentially noise-free projections. Under the verifiable assumption that we were working in the low-contrast limit, the lesion projections were then scaled down to those of a 6:1 lesion and added to the projections of the background. The 6:1 lesion-to-background ratio was chosen based on studies performed of actual healthy breast and malignant tissue after excisional biopsies following injection of Tc-99m-sestamibi, which found that typical *in vivo* tumor-background concentration ratios of Tc-99m-sestamibi are 5.64±3.06 [23].

The projection images were acquired by placing the phantom at the center of rotation of a Picker XP2000 two-headed SPECT system with the heads rotating at the minimum radius of rotation (9.0 cm). In this configuration the heads were within 2.0 cm of the walls of the phantom. The breast phantom was not attached to an anthropomorphic torso phantom because Wang *et al.* [22] showed that with proper shielding the contribution of scatter from the torso can be made negligible. We acquired 120 projections over 360° with each head acquiring to a 128x128 matrix (pixel size=4.67 mm). We used a low-energy, ultra-high resolution collimator.

The sinogram corresponding to the slice passing through the center of the lesion was extracted from the dataset. It was subsampled to generate sinograms of 15, 30, and 60 projections, as well as the complete 120-projection one. Each of these four sinograms was then reconstructed by use of four different methods:

1. Fourier-domain Hanning filtration of the projections (cutoff = 0.8 of Nyquist) followed by reconstruction from available views by FBP using an unapodized ramp filter.

2. Fourier-domain Hanning filtration of the projections (cutoff = 0.8), interpolation to 120 projections by use of ZP interpolation, and reconstruction by FBP using an unapodized ramp filter.

3. Spline-based, roughness-penalized Poisson likelihood smoothing of the projections, with a square root link function and the smoothing parameter λ chosen to give comparable smoothing to the Hanning filtration above, followed by reconstruction from available views by FBP using an unapodized ramp filter.

4. Spline-based, roughness-penalized Poisson likelihood smoothing of the projections, with a square root link function and the smoothing parameter λ chosen to give comparable smoothing to the Hanning filtration above, interpolation to 120 projections by use of ZP interpolation, and reconstruction by FBP using an unapodized ramp filter.

Naturally, no interpolation to 120 projections was performed for the sinograms already containing 120 projections.

2.8.2 Results—Reconstructing images for few-view SPECT SMM

The results of these reconstructions are shown in Fig. 4. In general, the star artifacts typical of few-view FBP reconstructions are evident in the reconstructions from 15 and 30 views, as is their suppression when these sinograms are interpolated to 120 views. Some mild interpolation artifacts are evident in the interpolated reconstructions from 15 views, but these do not obscure the lesion. No striking visual differences exist between the images reconstructed by use of the Fourier-based and spline-based approaches, but none were expected. These images are shown simply to demonstrate that both approaches produce reasonably satisfying images with suppressed noise and star artifacts without introducing significant new artifacts. A more quantitative comparison between the approaches will be presented in the next section.
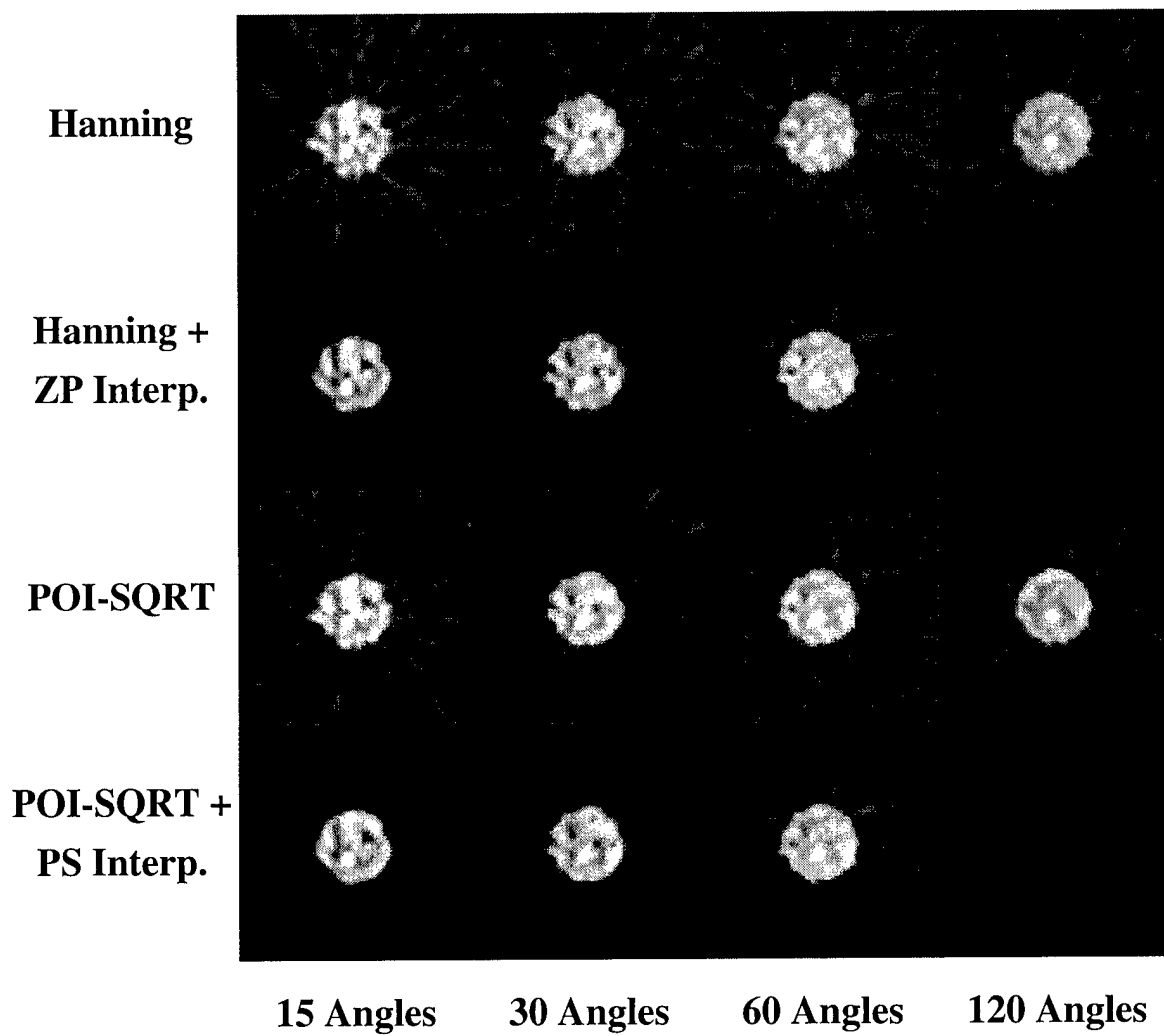
20

Figure 4: Reconstructions of a physical breast phantom containing a 6:1, 1-cm lesion imaged in a simulated dedicated SPECT SMM geometry.

2.8.3 Methods—Resolution-variance tradeoffs for few-view SPECT SMM

To characterize more quantitatively the differences between the Fourier-based and spline-based approaches to few-view SPECT SMM, we performed resolution-variance studies similar to those reported in the attached paper entitled, "Nonparametric regression sinogram smoothing using a roughness-penalized Poisson likelihood objective function." For this study, we employed a numerical phantom comprising a simple circle of uniform activity 1.0 specified on a 128×128 grid. Sinograms of 128 bins and 120 angles were generated, using strip integrals the same width as the bins, which in turn were the same size as the pixels. Similar sinograms of the same phantom containing an impulse of magnitude 0.01 in one pixel were also generated. The mean number of counts per bin was approximately 6–10. Each of these sinograms was subsampled to generate sinograms of 15, 30, and 60 projections, in addition to the original 120-projection sinograms.

To characterize resolution properties, the local impulse response (LIR, which is a generalization of the point spread function concept for images with potentially nonuniform resolution) for the pixel to which the impulse was added was computed for the four smoothing and interpolation methods enumerated in the previous sections. This was done for each number of starting angles: 15, 30, 60, and 120. The locally linear approximation discussed in the attached paper "Nonparametric regression sinogram smoothing using a roughness-penalized Poisson likelihood objective function," allowed the LIR for each method to be computed from the reconstructions of noiseless projections of the phantom with and without the impulse. The average FWHMs of each resulting LIR was computed by calculating the average diameter of the half-maximum contour of the LIR.

To characterize noise properties, 800 Poison realizations of the impulse-free projections were generated and reconstructed by each of the methods for each number of starting projections. The empirical standard deviation at the point of the impulse was computed for each method and number of starting projections.

2.8.4 Results—Resolution-variance tradeoffs for few-view SPECT SMM

The resulting resolution variance-curves are shown in Fig. 5. Two plots are shown there, one for the approaches in which smoothing alone is performed followed by reconstruction from the available number of projections, the other in which additional projections are interpolated after smoothing. Note that the ranges of the axes are not the same in both plots The solid lines in the center of each band represent the curves themselves and the shaded bands represent the 95% confidence intervals.

It can be seen that the spline-based approaches, both with and without interpolation of additional views, generally retain their superiority over the Fourier-based approaches in the few-view ECT context. A few other features of the curves bear mention. First, for all approaches, reconstructions from a smaller number of projections always have higher noise levels for a given resolution than reconstructions from a larger number of projections. This arises, of course, because fewer counts are contributing to the reconstructed image when fewer views are used for reconstruction. In fact, the relative noise levels are, as expected, more or less inversely proportional to the square root of the number of projections used. Second, the differences between the approaches are generally more apparent in the lower right portion of the plots than in the upper left portion. As in the paper "Nonparametric regression sinogram smoothing using a roughness-penalized Poisson likelihood objective function," this is because the upper left region of the plots correspond to relatively small values of the smoothing parameter and thus to images in which the two smoothing constraints have had little influence. Finally, and again not surprisingly, it should be noted that comparing corresponding curves between the two plots indicates that the projection interpolation process compromises image resolution somewhat, especially for small numbers of starting projections such as 15.

Technical Objective 5

To test the performance of the algorithms in a clinical setting, with the cooperation of Dr. Patrick Peller of Lutheran General Hospital in Chicago, who has proposed to develop a dedicated breast imaging miniature gamma camera in collaboration with the University of Chicago.
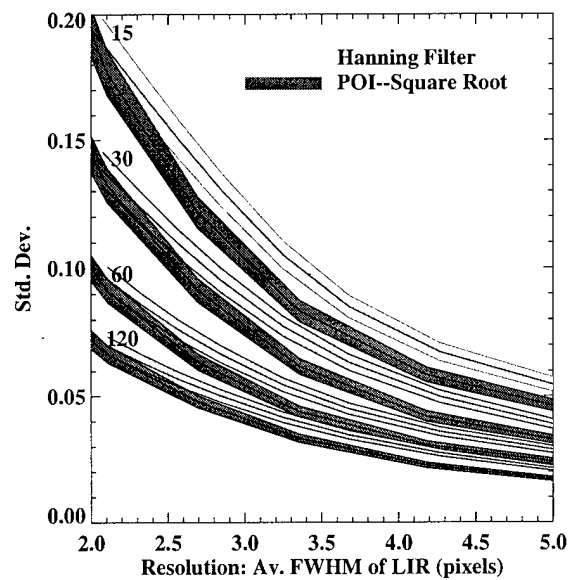
(a) Without interpolation of additonal projections.
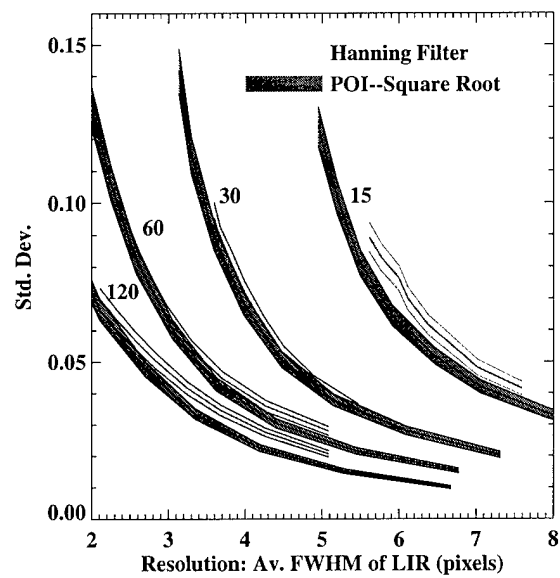
(b) With interpolation to 120 projections.

Figure 5: Resolution-variance curves for the Fourier- and spline-based projection smoothing and interpolation approaches in a breast-like phantom.

23

2.9 Tasks 9 and 10—Application of approaches to clinical SPECT SMM data

Obtain projections using the purpose-built breast mini-camera in a clinical situation. Reconstruct the distributions of MIBI using the few-view algorithms.

Evaluate the performance of the algorithms by comparing the diagnoses obtained from the receiver operating characteristic (ROC) analysis to determine the sensitivity and specificity of the technique.

As discussed in the introduction to this body section of the report, it was not possible to accomplish technical objective 5, and its subordinate tasks 9 and 10, because the clinical SPECT SMM system referred to was never developed. The development of the system was a project entirely independent of the work being reported on here, and it simply did not come to fruition due to lack of funding and personnel resources. In the future, it may be possible to evaluate the techniques developed in this work by seeking collaborations with groups that are currently developing miniature cameras specifically for the SMM application. These include a group at Lawrence Berkeley National Lab, lead by Bill Moses, and a group at the University of Rome, lead by R. Pani. However, neither of these systems is being developed with an explicitly tomographic capability, and thus a dedicated SPECT SMM system may still lie a few years in the future.

Key research accomplishments

- We have examined the use of the algebraic reconstruction technique algorithm for reconstruction of single photon emission computed tomography (SPECT) scintimammography (SMM) from a smaller number of projections than is usually used and found it to yield sufficiently poor results to preclude further examination of ART-type algorithms.

- We have analyzed the angular sampling requirements for the sinogram of a typical SPECT SMM slice and found that it is often the case that the Nyquist condition is satisfied by a smaller number of views than is needed by FBP to produce an essentially artifact-free reconstruction.

- From this analysis we concluded that it might be possible to interpolate additional views in a few-view SPECT SMM sinogram, thereby eliminating the source of star artifacts in few-view FBP reconstruction, without introducing additional artifacts or inaccuracies.

- We have developed a Fourier-based interpolation technique for increasing the number of projection views in few-view tomography with the aim of eliminating FBP's few-view artifacts.

- We have developed a spline-based interpolation technique for increasing the number of projection views in few-view tomography with the aim of eliminating FBP's few-view artifacts.

- We have compared the accuracy of Fourier, spline, and linear interpolation for increasing the number of projection views in few-view tomography and found spline interpolation to be, in general, the most accurate.

- We have analyzed the noise properties of Fourier-based, spline-based, and linear interpolation, and found that Fourier-based interpolation produces curves with the most favorably stationary noise properties, closely followed by spline-based interpolation.

- As an alternative to reconstruction by FBP, we have investigated the use of direct Fourier reconstruction, in which ZP interpolation is used to increase the density of polar samples in the Fourier transform space of the object being imaged, after which linear interpolation is used to estimate Cartesian samples for reconstruction by the 2D FFT. We have found this approach to be both accurate and computationally efficient.

- We have developed a novel Bayesian projection-smoothing technique based on roughness-penalized nonparametric regression using an explicit Poisson noise model.

- We have compared this novel projection-smoothing technique to conventional Fourier-domain apodization window techniques by computing resolution-variance curves and found the novel approach to be superior in a statistically significant way.

- We have found that the nonuniform resolution induced in the images by use of the novel projection smoothing technique can be controlled by adjusting the so-called *link function* of the model to yield the favorable outcomes of having better resolution in higher-count areas than in lower-count areas or of having essentially uniform resolution. This is in contrast to fully iterative Bayesian approaches, in which it is quite difficult to achieve uniform resolution.

- As a second alternative to FBP, for use when smoothing splines are fit to the projections, we have investigated a direct-spline reconstruction technique in which the reconstructed image is expressed in terms of the coefficients of the splines. The technique is found to yield image resolution superior to that of FBP, but at considerable computational cost.

- We have determined, through use of the ideal-observer framework, that the dedicated breast SPECT geometry is superior to the planar or conventional SPECT geometries for scintimammography.

- We have compared, using real breast phantom data, the algorithms based on the use of spline-based projection smoothing and interpolation to those based on Fourier-domain techniques. We found quantitative evidence of the superiority of the spline-based approach.

- Having found the algorithm based on the use of spline-based projection smoothing and interpolation to be superior to that based on Fourier-domain techniques, we determined that it can indeed preserve the diagnostic information that is present in reconstructions from standard numbers of projections when reconstructing from as few as 15 views for relatively simple objects such as breasts.

Reportable outcomes

Peer-Reviewed Original Articles

1. La Rivière, P. J. and Pan, X., "Nonparametric regression sinogram smoothing using a roughness-penalized Poisson likelihood objective function," accepted for publication in *IEEE Trans. Medical Imaging,* 2000.

2. La Rivière, P. J. and Pan, X., "Noise properties of periodic interpolation methods with implications for few-view tomography," *IEEE Trans. Nucl. Sci.,* 1999, **46,** pp. 639–645, 1999.

3. La Rivière, P. J. and Pan, X., "Few-view tomography using roughness penalized nonparametric regression and periodic spline interpolation," *IEEE Trans. Nucl. Sci.,* **46,** pp. 1121-1128, 1999.

4. Kao, C.-M., Pan, X., Anastasio, M., and La Rivière, P. J., "A Fourier-based optimal recovery approach for anti-aliasing interpolation," *Optical Engineering,* **38,** pp. 2041-2044, 1999.

5. La Rivière, P. J., Pan, X., and Penney, B. C., "Ideal-observer analysis of lesion detectability in planar, conventional SPECT, and dedicated SPECT scintimammography using effective multi-dimensional smoothing," *IEEE Trans. Nucl. Sci.,* **45,** pp. 1273-1279, 1998.

6. La Rivière, P. J. and Pan, X. "Spline-based inverse Radon transform in two and three dimensions," *IEEE Trans. Nucl. Sci.,* **45,** pp. 2224-2231, 1998.

Proceedings Articles

1. La Rivière, P. J. and Pan, X., "Resolution properties of non-parametric regression sinogram smoothing using an explicit Poisson model," *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.,* 1999 (CD-ROM).

2. La Rivière, P. J., Pan, X., and Kao, C.-M., "Medical imaging applications of a novel multi-dimensional interpolation approach," *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.,* 1999 (CD-ROM).

3. La Rivière, P. J. and Pan, X., "Few-view tomography using interpolating and smoothing splines with implications for cardiac SPECT," *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.,* **3,** pp. 1615-1619, 1999.

4. La Rivière, P. J. and Pan, X., "Noise properties of periodic interpolation methods with implications for few-view tomography," *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.,* **3,** pp. 1610-1614, 1999.

5. La Rivière, P. J. and Pan, X., "Comparison of angular interpolation approaches in few-view tomography using statistical hypothesis testing," *Proc. SPIE,* **3661,** pp. 398–407, 1999.

6. La Rivière, P. J., Pan, X., Penney, B. C., and C.-T., Chen, "Improved detectability of malignant lesions in SPECT scintimammography using effective multi-dimensional smoothing," *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.,* **2,** pp. 1581-1585, 1998.

7. La Rivière, P. J. and Pan, X., "Direct spline-based inversion of the three-dimensional Radon transform with application to cardiac phantom data," *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.,* **2,** pp. 1674-1679, 1998.

8. La Rivière, P. J. and Pan, X., "Mathematical equivalence of zero-padding interpolation and circular sampling theorem interpolation with implications for direct Fourier image reconstruction," *Proc. SPIE,* **3338,** pp. 1117-1126, 1998.

9. Pan, X., La Rivière, P. J., Ye, J., Mukherjee, J. and Chen, C.-T., "Efficient sinogram smoothing for dynamic neuroreceptor PET imaging," *Proc. SPIE,* **3033,** pp. 140-146, 1997.

Presentations

1. La Rivière, P. J. and Pan, X., "Resolution properties of non-parametric regression sinogram smoothing using an explicit Poisson model," *IEEE Medical Imaging Conference,* 1999.

2. La Rivière, P. J., Pan, X., and Kao, C.-M., "Medical imaging applications of a novel multi-dimensional interpolation approach," *IEEE Medical Imaging Conference,* 1999.

3. La Rivière, P. J. and Pan, X., "Comparison of angular interpolation approaches for few-view tomography using statistical hypothesis testing," *SPIE Medical Imaging Conference,* 1999.

4. La Rivière, P. J. and Pan, X., "Noise properties of periodic interpolation methods with implications for few-view tomography," *IEEE Medical Imaging Conference,* 1998.

5. Kao, C.-M., Pan, X., Anastasio, M., and La Rivière, P. J., "An interpolation method using signal recovery and discrete Fourier transform," *IEEE Medical Imaging Conference,* 1998.

6. La Rivière, P. J., Pan, X., and Penney B. C., "Few-view tomography using interpolating and smoothing splines with implications for cardiac SPECT," *IEEE Medical Imaging Conference,* 1998.

7. La Rivière, P. J. and Pan, X., "Mathematical equivalence of zero-padding interpolation and circular sampling theorem interpolation with implications for direct Fourier image reconstruction," *SPIE Medical Imaging Conference,* 1998.

8. La Rivière, P. J. and Pan, X., "Direct spline-based inversion of the three-dimensional Radon transform with application to cardiac phantom data," *IEEE Medical Imaging Conference,* 1997.

9. La Rivière, P. J. and Pan, X., "Improved detectability of malignant lesions in SPECT scintimammography using effective multi-dimensional smoothing, *IEEE Medical Imaging Conference,* 1997.

10. La Rivière, P. J., Pan, X., Penney, B. C., Chen, C.-T., and Reba, R., "Few-angle cardiac SPECT with spline processing: initial results and radial-extent limitations," *83rd Scientific Assembly and Annual Meeting of Radiological Society of North America,* 1997.

11. La Rivière, P. J. and Pan, X., "Direct spline-based inversion of the three-dimensional Radon transform with application to cardiac SPECT," *83rd Scientific Assembly and Annual Meeting of Radiological Society of North America,* 1997.

12. La Rivière, P. J., Pan, X., Penney, B. C., and Chen, C.-T., "Improved detectability of malignant lesions in dedicated SPECT scintimammography using effective multi-dimensional smoothing," *83rd Scientific Assembly and Annual Meeting of Radiological Society of North America,* 1997.

Degrees obtained

Patrick La Rivière expects to receive his Ph.D. in Radiology from the University of Chicago in December, 2000.

Conclusions

The broad objective of the proposed research was to develop, implement, and evaluate methods for the reconstruction of dedicated single-photon emission computed tomography (SPECT) scintimammography (SMM) images from a relatively small number of projection views. The research was carried out from a theoretical and abstract point of view, because no dedicated SPECT SMM system yet exists, and, indeed, one of the more significant conclusions of the work pertains to the more fundamental question of whether dedicated SPECT SMM is worth pursuing at all.

Through application of so-called "ideal observer" analysis to real data acquired using a simulated breast phantom, it was found that a dedicated SPECT SMM geometry should improve the detectability of focal lesions relative to planar imaging and conventional SPECT SMM imaging. Planar imaging is the current clinical standard for SMM imaging and conventional SPECT SMM imaging has been employed in occasional research studies without great success. The success of the dedicated SPECT geometry was attributed to the fact that it combines the advantages of the other two approaches: because of its small radius of rotation, it offers sensitivity and resolution comparable to that of planar imaging with the gamma camera flush against the breast, while it also offers the improved contrast offered by a tomographic system's ability to separate lesion activity from overlying and underlying activity. Of course, there are numerous issues that would have to be addressed before more final conclusions about the viability of dedicated breast SPECT systems could be drawn. These include technical questions about whether a system could be designed to image sufficiently close to the chest wall as well as practical questions about whether such a system would gain sufficient use to justify the purchase of so specialized a nuclear medicine system. The gamma cameras in the current planar SMM studies can be used for a wide range of other studies, and even the dedicated miniature breast gamma cameras currently under development are likely to be sufficiently flexible to permit application to other tasks, such as thyroid imaging. Depending on the specifics of its design, a dedicated SPECT SMM system would likely find few alternative applications. Nonetheless, as the first miniature dedicated miniature breast gamma cameras come to fruition, it will be worthwhile to study their performance in a tomographic geometry and thereby to continue refining the answer to this question.

Another practical issue that might plague dedicated SPECT SMM is the question of whether the imaging time needed to acquire the standard number of tomographic views and counts would be excessive, leading to motion artifacts and decreased throughput. In this vein, the driving motivation of this research was to develop techniques for the reconstruction of dedicated SPECT SMM images from a smaller number of views than is usually used for tomographic reconstruction. The research again began by addressing the more fundamental question of whether such few-view reconstruction is possible or advisable in this situation. Analysis of angular sampling requirements for sinograms of breast-like phantoms suggested that the Nyquist sampling condition could be satisfied by a relatively small number of views in SPECT SMM and thus that reconstruction should be possible from small numbers of views without significant angular aliasing artifacts. This should not, of course, be taken to mean that few-view acquisitions are the only or even necessarily the best means of reducing imaging time in dedicated SPECT SMM. The same time savings could, of course, be achieved by acquiring fewer counts at each of a standard number of views. The question of which of these approaches is best was not a subject of this investigation, although interestingly, the Bayesian projection-smoothing approach developed in the course of this research could be of use for smoothing the noisy, low-count projections acquired in the latter approach.

The analysis of angular sampling requirements in SPECT SMM, coupled with an analysis of the artifacts that arise in few-view filtered backprojection (FBP) reconstruction, lead to the development of a sinogram preprocessing approach for few-view SPECT SMM reconstruction, in which each measured projection is first smoothed by use of a spline-based, Bayesian, nonparametric regression technique that explicitly models the Poisson statistics of the measurement data. Additional projections are then interpolated from the measured ones in the hopes of satisfying FBP's own assumptions about the completeness of the measured data. Reconstruction then proceeds by use of FBP. The Bayesian projection-smoothing approach has a number of appealing features: it leads to statistically significant resolution/noise advantages over Fourier-domain apodization windows, it captures some of the statistical benefits of a fully iterative reconstruction algorithm, such as MLEM, at a fraction of the computational cost, and it also can also

29

be used to produce images having uniform and isotropic resolution more naturally than when making use of fully iterative algorithms. Significantly, this Bayesian projection smoothing approach could be applied quite generally in emission tomography and is not tied specifically to the few-view application or to SPECT SMM. The investigation of interpolation methods for generating additional views in a few-view sinogram also produced insights and methods of broad significance, as it lead to a detailed investigation of the noise and accuracy properties of various interpolation methods. Overall, spline-based interpolation was found to perform best in the SPECT SMM context. The application of the spline-based smoothing and interpolation approaches in concert to few-view SPECT SMM data was found to yield visually satisfying reconstructions from as few as 15 projection views, although in practice 30 may be a preferable lower limit. That such reconstructions still retain diagnostically important information was confirmed by imaging a cardiac phantom containing a defect insert and applying to the reconstructed data standard analysis tools that are very sensitive to the presence of such defects in reconstructed images.

So what?

The value of the accomplished research lies in three principal elements. First, it has established that a dedicated breast SPECT geometry for SMM imaging should lead to enhanced lesion detectability relative to clinically standard planar imaging and conventional SPECT imaging. This suggests that the development of dedicated SPECT SMM systems should be considered in the future as they may lead to improvements in the sensitivity of SMM imaging and thereby make it more viable as a follow-up test to mammography that attempts that attempts to differentiate benign and malignant lesions noninvasively. For dedicated SPECT SMM to be a truly viable follow-up to mammography, however, the test should be relatively comfortable and its throughput relatively high. In this vein, the second significant element of the research was its finding that it should, in principle, be possible to reconstruct dedicated SPECT SMM images from a smaller number of views than is usually used, without increasing the time spent acquiring each such view, and thus to keep imaging time relatively short. The final significant aspect of the research was the development of methods for performing this kind of few-view reconstruction efficiently and accurately. In addition, some of the techniques and insights developed here will likely have resonance beyond the SPECT SMM and few-view reconstruction problems and find application to other problems in emission tomography and in medical imaging in general.

References

[1] G. T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography.* Boston: Academic Press, 1980.

[2] R. L. Siddon, "Fast calculation of the exact radiological path for a three-dimensional CT array," *Med. Phys.,* vol. 12, pp. 252–255, 1985.

[3] R. A. Brooks, G. H. Weiss, and A. J. Talbert, "A new approach to interpolation in computed tomography," *J. Comput. Assist. Tomogr.,* vol. 2, pp. 577–585, 1978.

[4] H. Stark, "Sampling theorems in polar coordinates," *J. Opt. Soc. Am.,* vol. 69, pp. 1519–1525, 1979.

[5] H. Stark and M. Wengrovitz, "Comments and corrections on the use of polar sampling theorems in CT," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 31, pp. 1329–1331, 1983.

[6] R. W. Schafer and L. R. Rabiner, "A digital signal processing approach to interpolation," *Proc. IEEE,* vol. 61, pp. 692–702, 1973.

[7] K. P. Prasad and P. Satyanarayana, "Fast interpolation algorithm using FFT," *Electron. Lett.,* vol. 22, pp. 185–187, 1986.

[8] D. Fraser, "Interpolation by the FFT revisited—an experimental investigation," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 37, pp. 665–675, 1989.

[9] J. A. Fessler, "Tomographic reconstruction using information-weighted spline smoothing," in *Information Processing in Medical Imaging* (H. H. Barrett and A. F. Gmitro, eds.), pp. 290–300, Berlin: Springer-Verlag, 1993.

[10] P. J. Green and B. W. Silverman, *Nonparametric Regression and Generalized Linear Models.* London: Chapman Hall, 1994.

[11] H. Stark, J. W. Woods, I. Paul, and R. Hingorani, "Direct Fourier reconstruction in computer tomography," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 29, pp. 237–245, 1981.

[12] H. Stark, J. W. Woods, I. Paul, and R. Hingorani, "An investigation of computerized tomography by direct Fourier inversion and optimum interpolation," *IEEE Trans. Biomed. Eng.,* vol. 28, pp. 496–505, 1981.

[13] S. X. Pan and A. C. Kak, "A computational study of reconstruction algorithms for diffraction tomography: Interpolation versus filtered backpropagation," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. 31, pp. 1262–1275, 1983.

[14] P. J. La Rivière and X. Pan, "Comparison of angular interpolation approaches in few-view tomography using statistical hypothesis testing," in *Proc. SPIE,* vol. 3661, pp. 398–407, 1999.

[15] International Comission on Radiation Units and Measurements, "Medical imaging: The assessment of image quality," 7910 Woodmont Ave., Bethesda, MD, 20814, 1996.

[16] S. C. Moore, M. F. Kijewski, S. P. Müller, and B. L. Holman, "SPECT image noise power: Effects of nonstationary projection noise and attenuation compensation," *J. Nucl. Med.,* vol. 29, pp. 1704–1709, 1988.

[17] E. V. Garcia, K. V. Train, J. Maddahi, F. Prigent, J. Friedman, J. Areeda, A. Waxman, and D. S. Berman, "Quantification of rotational thallium-201 myocardial tomography," *J. Nucl. Med.,* vol. 26, pp. 17–26, 1985.

[18] P. J. La Rivière and X. Pan, "Few-view tomography using roughness-penalized nonparametric regression and periodic spline interpolation," *IEEE Trans. Nucl. Sci.,* vol. 46, pp. 1121–1128, 1999.

[19] F. O'Sullivan, B. S. Yandell, and W. J. Raynor, Jr., "Automatic smoothing of regression functions in generalized linear models," *J. Amer. Stat. Assoc.*, vol. 81, pp. 96–103, 1986.

[20] C. Gu, "Adaptive spline smoothing in non gaussian regression models," *J. Amer. Stat. Assoc.*, vol. 85, pp. 801–807, 1990.

[21] C. Gu, "Cross-validating non-gaussian data," *J. of Computational and Graphical Statistics*, vol. 1, pp. 169–179, 1992.

[22] H. Wang, C. Scarfone, K. L. Greer, R. E. Coleman, and R. J. Jaszczak, "Prone breast tumor imaging using vertical axis-of-rotation SPECT systems: An initial study," *IEEE Trans. Nucl. Sci.*, vol. 44, pp. 1271–1276, 1997.

[23] J. Maublant, M. de Latour, D. Mestas, A. Clemenson, S. Charrier, V. Feillel, G. Le Bouedec, P. Kaufmann, J. Dauplat, and A. Veyre, "Technetium-99m-sestamibi uptake in breast tumor and associated lymph nodes," *J. Nucl. Med.*, vol. 37, pp. 922–925, 1996.

Final report bibliography and personnel

Peer-Reviewed Original Articles

1. La Rivière, P. J. and Pan, X., "Nonparametric regression sinogram smoothing using a roughness-penalized Poisson likelihood objective function," accepted for publication in *IEEE Trans. Medical Imaging*, 2000.

2. La Rivière, P. J. and Pan, X., "Noise properties of periodic interpolation methods with implications for few-view tomography," *IEEE Trans. Nucl. Sci.*, 1999, **46**, pp. 639–645, 1999.

3. La Rivière, P. J. and Pan, X., "Few-view tomography using roughness penalized nonparametric regression and periodic spline interpolation," *IEEE Trans. Nucl. Sci.*, **46**, pp. 1121-1128, 1999.

4. Kao, C.-M., Pan, X., Anastasio, M., and La Rivière, P. J., "A Fourier-based optimal recovery approach for anti-aliasing interpolation," *Optical Engineering*, **38**, pp. 2041-2044, 1999.

5. La Rivière, P. J., Pan, X., and Penney, B. C., "Ideal-observer analysis of lesion detectability in planar, conventional SPECT, and dedicated SPECT scintimammography using effective multi-dimensional smoothing," *IEEE Trans. Nucl. Sci.*, **45**, pp. 1273-1279, 1998.

6. La Rivière, P. J. and Pan, X. "Spline-based inverse Radon transform in two and three dimensions," *IEEE Trans. Nucl. Sci.*, **45**, pp. 2224-2231, 1998.

Proceedings Articles

1. La Rivière, P. J. and Pan, X., "Resolution properties of non-parametric regression sinogram smoothing using an explicit Poisson model," *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 1999 (CD-ROM).

2. La Rivière, P. J., Pan, X., and Kao, C.-M., "Medical imaging applications of a novel multi-dimensional interpolation approach," *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 1999 (CD-ROM).

3. La Rivière, P. J. and Pan, X., "Few-view tomography using interpolating and smoothing splines with implications for cardiac SPECT," *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, **3**, pp. 1615-1619, 1999.

4. La Rivière, P. J. and Pan, X., "Noise properties of periodic interpolation methods with implications for few-view tomography," *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, **3**, pp. 1610-1614, 1999.

5. La Rivière, P. J. and Pan, X., "Comparison of angular interpolation approaches in few-view tomography using statistical hypothesis testing," *Proc. SPIE*, **3661**, pp. 398–407, 1999.

6. La Rivière, P. J., Pan, X., Penney, B. C., and C.-T., Chen, "Improved detectability of malignant lesions in SPECT scintimammography using effective multi-dimensional smoothing," *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, **2**, pp. 1581-1585, 1998.

7. La Rivière, P. J. and Pan, X., "Direct spline-based inversion of the three-dimensional Radon transform with application to cardiac phantom data," *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, **2**, pp. 1674-1679, 1998.

8. La Rivière, P. J. and Pan, X., "Mathematical equivalence of zero-padding interpolation and circular sampling theorem interpolation with implications for direct Fourier image reconstruction," *Proc. SPIE*, **3338**, pp. 1117-1126, 1998.

9. Pan, X., La Rivière, P. J., Ye, J., Mukherjee, J. and Chen, C.-T., "Efficient sinogram smoothing for dynamic neuroreceptor PET imaging," *Proc. SPIE*, **3033**, pp. 140-146, 1997.

Presentations

1. La Rivière, P. J. and Pan, X., "Resolution properties of non-parametric regression sinogram smoothing using an explicit Poisson model," *IEEE Medical Imaging Conference,* 1999.

2. La Rivière, P. J., Pan, X., and Kao, C.-M., "Medical imaging applications of a novel multi-dimensional interpolation approach," *IEEE Medical Imaging Conference,* 1999.

3. La Rivière, P. J. and Pan, X., "Comparison of angular interpolation approaches for few-view tomography using statistical hypothesis testing," *SPIE Medical Imaging Conference,* 1999.

4. La Rivière, P. J. and Pan, X., "Noise properties of periodic interpolation methods with implications for few-view tomography," *IEEE Medical Imaging Conference,* 1998.

5. Kao, C.-M., Pan, X., Anastasio, M., and La Rivière, P. J., "An interpolation method using signal recovery and discrete Fourier transform," *IEEE Medical Imaging Conference,* 1998.

6. La Rivière, P. J., Pan, X., and Penney B. C., "Few-view tomography using interpolating and smoothing splines with implications for cardiac SPECT," *IEEE Medical Imaging Conference,* 1998.

7. La Rivière, P. J. and Pan, X., "Mathematical equivalence of zero-padding interpolation and circular sampling theorem interpolation with implications for direct Fourier image reconstruction," *SPIE Medical Imaging Conference,* 1998.

8. La Rivière, P. J. and Pan, X., "Direct spline-based inversion of the three-dimensional Radon transform with application to cardiac phantom data," *IEEE Medical Imaging Conference,* 1997.

9. La Rivière, P. J. and Pan, X., "Improved detectability of malignant lesions in SPECT scintimammography using effective multi-dimensional smoothing, *IEEE Medical Imaging Conference,* 1997.

10. La Rivière, P. J., Pan, X., Penney, B. C., Chen, C.-T., and Reba, R., "Few-angle cardiac SPECT with spline processing: initial results and radial-extent limitations," *83rd Scientific Assembly and Annual Meeting of Radiological Society of North America,* 1997.

11. La Rivière, P. J. and Pan, X., "Direct spline-based inversion of the three-dimensional Radon transform with application to cardiac SPECT," *83rd Scientific Assembly and Annual Meeting of Radiological Society of North America,* 1997.

12. La Rivière, P. J., Pan, X., Penney, B. C., and Chen, C.-T., "Improved detectability of malignant lesions in dedicated SPECT scintimammography using effective multi-dimensional smoothing," *83rd Scientific Assembly and Annual Meeting of Radiological Society of North America,* 1997.

Personnel receiving pay

1. Patrick La Rivière

34

Appendices : Attached articles

The following articles, which provide supporting material and data for this report, are appended in the order in which they are cited in the text:

1. La Rivière, P. J. and Pan, X., "Mathematical equivalence of zero-padding interpolation and circular sampling theorem interpolation with implications for direct Fourier image reconstruction," *Proc. SPIE, 3338,* pp. 1117-1126, 1998.

2. La Rivière, P. J. and Pan, X., "Comparison of angular interpolation approaches in few-view tomography using statistical hypothesis testing," *Proc. SPIE, 3661,* pp. 398–407, 1999.

3. La Rivière, P. J. and Pan, X., "Noise properties of periodic interpolation methods with implications for few-view tomography," *IEEE Trans. Nucl. Sci.,* 1999, **46**, pp. 639–645, 1999.

4. La Rivière, P. J. and Pan, X., "Few-view tomography using roughness penalized nonparametric regression and periodic spline interpolation," *IEEE Trans. Nucl. Sci.,* **46,** pp. 1121-1128, 1999.

5. La Rivière, P. J. and Pan, X., "Nonparametric regression sinogram smoothing using a roughness-penalized Poisson likelihood objective function," accepted for publication in *IEEE Trans. Medical Imaging,* 2000.

6. La Rivière, P. J. and Pan, X. "Spline-based inverse Radon transform in two and three dimensions," *IEEE Trans. Nucl. Sci.,* **45,** pp. 2224-2231, 1998.

7. La Rivière, P. J., Pan, X., and Penney, B. C., "Ideal-observer analysis of lesion detectability in planar, conventional SPECT, and dedicated SPECT scintimammography using effective multi-dimensional smoothing," *IEEE Trans. Nucl. Sci.,* **45**, pp. 1273-1279, 1998.

Mathematical equivalence of zero-padding interpolation and circular sampling theorem interpolation with implications for direct Fourier image reconstruction

P. J. La Rivière and X. Pan

Department of Radiology, The University of Chicago, Chicago, IL, 60637

ABSTRACT

The speed and accuracy of Direct Fourier image reconstruction methods have long been hampered by the need to interpolate between the polar grid of Fourier data that is obtained from the measured projection data and the Cartesian grid of Fourier data that is needed to recover an image using the 2D FFT. Fast but crude interpolation schemes such as bilinear interpolation often lead to unacceptable image artifacts, while more sophisticated but computationally intense techniques such as circular sampling theorem (CST) interpolation negate the speed advantages afforded by the use of the 2D FFT. One technique that has been found to yield high-quality images without much computational penalty is a hybrid one in which zero-padding interpolation is first used to increase the density of samples on the polar grid after which bilinear interpolation onto the Cartesian grid is performed. In this work, we attempt to account for the success of the approach relative to the CST approach in three ways. First and most importantly, we establish that zero-padding interpolation of periodic functions that are sampled in accordance with the Nyquist criterion—precisely the sort of function encountered in the angular dimension of the polar grid—is exact and equivalent to circular sampling theorem interpolation. Second, we point out that both approaches make comparable approximations in interpolating in the radial direction. Finally, we indicate that the error introduced by the bilinear interpolation step in the zero-padding approach can be minimized by choosing sufficiently large zero-padding factors.

Keywords: Interpolation, Image Reconstruction, Direct Fourier Methods, Zero-padding, Circular Sampling Theorem, FFT

1. INTRODUCTION

The direct Fourier approach to tomographic image reconstruction has long had the potential to be among the most computationally efficient of image reconstruction algorithms because it harnesses the speed of the inverse 2D Fast Fourier Transform (FFT). The inverse 2D FFT is used to recover an image from samples of its Fourier transform that are obtained by Fourier transforming the projections of the object slice being imaged. As is widely known, the chief impediment to effective direct Fourier algorithms is the fact that the transformed projection data yields samples of the image transform that lie on a polar grid, while the inverse 2D FFT algorithm requires samples to lie on a Cartesian grid. The necessary interpolation between the two grids generally compromises either image quality, when a fast but crude interpolation is performed, or processing time, when a more sophisticated but computationally intense interpolation is performed. For instance, the simplest of interpolation schemes—nearest neighbor or bilinear interpolation—can lead to artifacts in the reconstructed images [1], while more accurate techniques, such as the gridding approaches of O'Sullivan [2] and Jackson *et al.* [3], involve more expensive convolutions over larger neighborhoods.

Stark *et al.* [4,5] have examined an interpolation scheme that is essentially exact when the polar samples of the transform satisfy certain conditions. They use truncated Whittaker-Shannon sinc interpolation in the radial direction and circular sampling theorem (CST) interpolation in the angular direction. CST interpolation is an exact interpolation method that follows from Whittaker-Shannon interpolation when the function being interpolated is periodic and sampled in accordance with the Nyquist criterion. Unfortunately, CST interpolation also involves a very expensive convolution, a processing burden Stark *et al.* attempted to mitigate by truncating the CST series, with a consequent tradeoff in accuracy.

At least one interpolation approach has been developed that generates images of comparable quality to CST and other convolution-based techniques with none of their computational drawbacks. It is a hybrid method in which zero-padding interpolation is first used to increase the density of polar samples in both the radial and angular directions, after which bilinear interpolation onto the Cartesian grid is performed. Zero-padding interpolation works by extending the discrete Fourier transform (DFT) of a finite sequence with zeroes and then performing an inverse DFT, yielding a sequence with additional, interpolated samples between the measured values of the original sequence. Because it exploits the FFT algorithm, zero-padding interpolation is very computationally efficient. The approach was studied explicitly and compared to CST-based techniques by Kak and Pan in the context of ultrasound diffraction tomography, where a similar Fourier-space interpolation problem arises [6].

In this work, we account for the surprising success of this hybrid approach in three ways. First, we establish that for the sort of periodic, bandlimited functions encountered in the angular dimension of the polar grid, zero-padding interpolation is equivalent to CST interpolation insofar as the values zero-padding interpolation generates on the denser polar grid it produces match the values CST interpolation would produce at those points. Of course, CST interpolation is not constrained to interpolate onto a polar grid of increased angular density—it can interpolate at arbitrary angular points—and thus avoids the additional linear interpolation needed in the zero-padding approach. However, we point out that the error arising from this last interpolation can be made negligibly small if the polar grid is made sufficiently dense through zero-padding. Finally, we indicate that both the zero-padding and CST approaches make comparable approximations in interpolating in the radial direction of the polar grid.

We begin in part 2 by reviewing the basic theory of CST and zero-padding interpolation and then proving their equivalence for one-dimensional, periodic functions sampled in accordance with the Nyquist criterion. In part 3, we introduce the specific interpolation problem encountered in direct Fourier image reconstruction and discuss the solution offered by CST and zero-padding interpolation, as well as simple bilinear interpolation. In part 4, we present the results of using these three methods to reconstruct images of numerical phantoms, both ideal and contaminated by Poisson noise.

2. THEORY

In this section we focus on the interpolation of one-dimensional, bandlimited, periodic functions of period 2π, introducing the CST and zero-padding approaches and then establishing their agreement on the grid of zero-padding interpolated values. The decision to focus on functions of period 2π simplifies notation and reflects the fact that this is the period of the angular samples encountered in direct Fourier reconstruction, but all of the results derived apply equally well to functions with other periods.

2.1. Circular Sampling Theorem

Circular Sampling Theorem interpolation is a special case of exact Whittaker-Shannon interpolation that applies to periodic functions sampled in accordance with the Nyquist criterion [7,8]. Consider a function $x(\theta)$ that is periodic with period 2π and bandlimited to frequency K (i.e., the coefficients of the function's Fourier series expansion satisfy $a_k = 0$ for $|k| > K$). If there are N equidistant samples of $x(\theta)$, i.e., $x(2\pi n/N)$, $n = 0,...,N-1$, and N is odd and $\geq 2K+1$, the circular sampling theorem (CST) states that the value of $x(\theta)$ may be determined exactly at arbitrary θ using

$$x(\theta) = \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right) \frac{\sin\left[\frac{N}{2}\left(\theta - \frac{2\pi n}{N}\right)\right]}{N\sin\left[\frac{1}{2}\left(\theta - \frac{2\pi n}{N}\right)\right]}. \tag{1}$$

Similarly, if N is even and $\geq 2K$, the value of $x(\theta)$ may be determined exactly at arbitrary θ using

$$x(\theta) = \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right) \frac{\sin\left[\frac{1}{2}(N-1)\left(\theta - \frac{2\pi n}{N}\right)\right]}{N\sin\left[\frac{1}{2}\left(\theta - \frac{2\pi n}{N}\right)\right]}. \tag{2}$$

2.2. Zero-padding interpolation

It is well known that extending the discrete Fourier Transform (DFT) of a finite sequence with zeroes and performing an inverse DFT increases the density of samples in the conjugate domain, thereby yielding interpolated values at regular intervals between the sequence's measured points. Fraser has given a thorough and enlightening discussion of the process that highlights a few of its more subtle points [9]. It is not widely known, however, that interpolation by zero-padding the DFT is equivalent to CST interpolation (and thus to exact Whittaker-Shannon interpolation) when the sequence in question represents samples of a periodic, bandlimited function sampled in accordance with the Nyquist criterion. To establish this equivalence, we begin with a rigorous definition of the zero-padding process, taking care to distinguish between the cases when the number of samples N is even or odd. Following Fraser, we define a superscript $^-$, such that $N^- = N$ (N even) and $N^- = N-1$ (N odd). Assume again that we have a function $x(\theta)$ that is periodic with period 2π and bandlimited to frequency K, and of which we have N equidistant samples, $x(2\pi n/N)$, $n = 0,...,N-1$, where $N \geq 2K$. The DFT of this sequence is given by

$$c_k = \frac{1}{N}\sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\exp(-j2\pi nk / N), \quad k = 0,...,N-1, \tag{3}$$

where $j = \sqrt{-1}$. Zero-padding involves the creation of a new sequence $d_{k'}$, having $L = P \cdot N$ terms, where P is an integer. Taking the inverse DFT of the new sequence yields a more densely sampled version of the original sequence. We need only explicitly build the first half of the DFT sequence $d_{k'}$, for so long as x is a real function we can exploit the conjugate symmetry of the DFT to construct the second half. The first half is created as follows:

$$
d_{k'} = \begin{cases}
c_{k'} & k' = 0, 1, 2, \ldots, N^- / 2 - 1 \\
0.5 c_{k'} & k' = N^- / 2 \quad (N \text{ even}) \\
c_{k'} & k' = N^- / 2 \quad (N \text{ odd}) \\
0 & k' = N^- / 2 + 1, N^- / 2 + 2, \ldots, L^- / 2,
\end{cases}
\tag{4}
$$

where L^- is defined in the same way as N^- above. The second half of the sequence $d_{k'}$ may then be constructed from the first half by use of

$$
d_{L-k'} = d_{k'}^* \quad k = 1, 2, \ldots, L^- / 2,
\tag{5}
$$

where $*$ denotes complex conjugation. The factor of 1/2 that appears in Eq. (4) in front of $c_{N/2}$ when N is even and the implicit insertion of this term's complex conjugate in the second half of the sequence are often omitted in hasty implementations of zero-padding interpolation in which sequences of zeroes are simply shoehorned into the middle of the DFT sequence. Fraser discusses the mathematical reason for these nuances, which will prove to be important in establishing the mathematical equivalence of CST and zero-padding for the N even case. Finally then, taking the inverse DFT of the sequence $d_{k'}$,

$$
x\left(\frac{2\pi l}{L}\right) = \sum_{k'=0}^{L-1} d_{k'} \exp(j 2\pi l k' / L),
\tag{6}
$$

yields the more densely sampled sequence $x(2\pi l / L)$, $l = 0, \ldots, L - 1$.

2.3. Equivalence of Zero-Padding and CST Interpolation

The most obvious mechanical difference between CST interpolation and zero-padding interpolation is that the former can be used to interpolate a value at a single, arbitrary point θ while the latter necessarily generates simultaneously numerous interpolated points, which lie on a regular grid whose spacing is determined by the zero-padding factor P. However, we will now demonstrate that the values determined by zero-padding on that regular grid match exactly the values that would be obtained from CST interpolation at those points.

We consider the task of using zero-padding to interpolate from N samples, $x(2\pi n / N)$, $n = 0, \ldots, N - 1$, to L samples, $x(2\pi l / N)$, $l = 0, \ldots, L - 1$, where $L = P \cdot N$ as before. Consider first the case where N is odd. We can divide the sum in Eq. (6) into three segments:

$$
x\left(\frac{2\pi l}{L}\right) = \sum_{k'=0}^{(N-1)/2} d_{k'} \exp(j 2\pi l k' / L) + \sum_{k'=(N-1)/2+1}^{L-(N-1)/2-1} d_{k'} \exp(j 2\pi l k' / L) + \sum_{k'=L-(N-1)/2}^{L-1} d_{k'} \exp(j 2\pi l k' / L).
\tag{7}
$$

Substituting for the $d_{k'}$ in terms of the $c_{k'}$ as specified by Eq. (4), we have

$$
x\left(\frac{2\pi l}{L}\right) = \sum_{k'=0}^{(N-1)/2} c_{k'} \exp(j 2\pi l k' / L) + \sum_{k'=L-(N-1)/2}^{L-1} c_{L-k'}^* \exp(j 2\pi l k' / L),
\tag{8}
$$

where the second sum has disappeared because the coefficients $d_{k'}$ are all 0 for k' in this range. Substituting the expression for $c_{k'}$ given by Eq. (3) yields

$$
\begin{aligned}
x\left(\frac{2\pi l}{L}\right) = & \sum_{k'=0}^{(N-1)/2} \left[\frac{1}{N} \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right) \exp(-j 2\pi n k' / N) \right] \exp(j 2\pi l k' / L) \\
& + \sum_{k'=L-(N-1)/2}^{L-1} \left[\frac{1}{N} \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right) \exp(j 2\pi n (L - k') / N) \right] \exp(j 2\pi l k' / L).
\end{aligned}
\tag{9}
$$

Switching the order of the summations and making the substitution $k'' = k' - L$ in the second term gives

$$
x\left(\frac{2\pi l}{L}\right) = \frac{1}{N} \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right) \left[\sum_{k'=0}^{(N-1)/2} \exp(-j 2\pi n k' / N) \exp(j 2\pi l k' / L) + \sum_{k''=-(N-1)/2}^{-1} \exp(-j 2\pi n k'' / N) \exp(j 2\pi l (k'' + L) / L) \right].
\tag{10}
$$

The expression $\exp(j2\pi l(k''+L)/L)$ in the second term in brackets is simply equal to $\exp(j2\pi lk''/L)$, so the two terms in brackets may be combined to give

$$x\left(\frac{2\pi l}{L}\right) = \frac{1}{N}\sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\left[\sum_{k'=-(N-1)/2}^{(N-1)/2}\exp\left[-j2\pi k'\left(\frac{n}{N}-\frac{l}{L}\right)\right]\right]. \tag{11}$$

At this point we use an identity, which holds for N odd, also invoked by Stark in his derivation of the CST [7]:

$$\frac{1}{N}\sum_{k'=-(N-1)/2}^{(N-1)/2}\exp\left[-j2\pi k'\left(\frac{n}{N}-\frac{l}{L}\right)\right] = \frac{\sin\left[\frac{N}{2}\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}{N\sin\left[\frac{1}{2}\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}. \tag{12}$$

Using this in Eq. (11) yields

$$x\left(\frac{2\pi l}{L}\right) = \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\frac{\sin\left[\frac{N}{2}\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}{N\sin\left[\frac{1}{2}\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}. \tag{13}$$

We see that the values $x(2\pi l/N)$ obtained using zero-padding are exactly those that would be obtained from Eq. (1) for CST interpolation with N odd and $\theta = 2\pi l/L$.

When N is even, the derivation proceeds slightly differently. We now divide the sum in Eq. (6) into five segments:

$$x\left(\frac{2\pi l}{L}\right) = \sum_{k'=0}^{N/2-1} d_{k'}\exp(j2\pi lk'/L) + 0.5 d_{N/2}\exp(j2\pi lN/2L) + \sum_{k'=N/2+1}^{L-N/2-1} d_{k'}\exp(j2\pi lk'/L)$$
$$+0.5 d_{L-N/2}\exp(j2\pi l(L-N/2)/L) + \sum_{k'=L-N/2+1}^{L-1} d_{k'}\exp(j2\pi lk'/L). \tag{14}$$

Substituting for the $d_{k'}$ in terms of the $c_{k'}$ as specified by Eq. (4), we have

$$x\left(\frac{2\pi l}{L}\right) = \sum_{k'=0}^{N/2-1} c_{k'}\exp(j2\pi lk'/L) + 0.5 c_{N/2}\exp(j2\pi lN/2L)$$
$$+0.5 c_{N/2}^*\exp(j2\pi l(L-N/2)/L) + \sum_{k'=L-N/2+1}^{L-1} c_{L-k'}^*\exp(j2\pi lk'/L), \tag{15}$$

where the third term has disappeared because the coefficients $d_{k'}$ are all 0 for k' in this range. Substituting Eq. (3) for $c_{k'}$ yields

$$x\left(\frac{2\pi l}{L}\right) = \sum_{k'=0}^{N/2-1}\left[\frac{1}{N}\sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\exp(-j2\pi nk'/N)\right]\exp(j2\pi lk'/L)$$
$$+0.5\left[\frac{1}{N}\sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\exp(-j2\pi n(N/2)/N)\right]\exp(j2\pi lN/2L)$$
$$+0.5\left[\frac{1}{N}\sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\exp(j2\pi n(N/2)/N)\right]\exp(j2\pi l(L-N/2)/L)$$
$$+\sum_{k'=L-N/2+1}^{L-1}\left[\frac{1}{N}\sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\exp(j2\pi n(L-k')/N)\right]\exp(j2\pi lk'/L). \tag{16}$$

Now switching the order of the summations, making the substitution $k'' = k' - L$ in the fourth term, and reordering the terms for later convenience allows us to write

$$x\left(\frac{2\pi l}{L}\right) = \frac{1}{N}\sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\left[\sum_{k'=0}^{N/2-1}\exp(-j2\pi nk'/N)\exp(j2\pi lk'/L) + \sum_{k''=-(N/2-1)}^{-1}\exp(-j2\pi nk''/N)\exp(j2\pi l(k''+L)/L)\right.$$
$$\left. +0.5\exp(-j\pi n)\exp(j\pi lN/L) + 0.5\exp(j\pi n)\exp(-j\pi lN/L)\right]. \tag{17}$$

The factor $\exp(j2\pi l(k''+L)/L)$ in the second sum is simply equal to $\exp(j2\pi lk''/L)$, so the first two sums in brackets may be combined, as can the third and fourth terms, yielding

$$x\left(\frac{2\pi l}{L}\right) = \frac{1}{N}\sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\left[\sum_{k'=-(N/2-1)}^{N/2-1} \exp\left(-j2\pi k'\left(\frac{n}{N}-\frac{l}{L}\right)\right) + \cos\left[\pi N\left(\frac{l}{L}-\frac{n}{N}\right)\right]\right]. \tag{18}$$

We now invoke an identity that holds for N even and whose derivation is similar to that of Eq. (12) [8]:

$$\frac{1}{N}\sum_{k'=-(N/2-1)}^{N/2-1} \exp\left[-j2\pi k'\left(\frac{n}{N}-\frac{l}{L}\right)\right] = \frac{\sin\left[\frac{1}{2}(N-1)\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}{N\sin\left[\frac{1}{2}\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}. \tag{19}$$

This allows us to write

$$x\left(\frac{2\pi l}{L}\right) = \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\left[\frac{\sin\left[\frac{1}{2}(N-1)\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}{N\sin\left[\frac{1}{2}\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]} + \cos\left[\pi N\left(\frac{l}{L}-\frac{n}{N}\right)\right]\right]. \tag{20}$$

Stark and Wengrovitz point out that the second term contributes 0 to $x(2\pi l/N)$, so the final expression is given by

$$x\left(\frac{2\pi l}{L}\right) = \sum_{n=0}^{N-1} x\left(\frac{2\pi n}{N}\right)\left[\frac{\sin\left[\frac{1}{2}(N-1)\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}{N\sin\left[\frac{1}{2}\left(\frac{2\pi l}{L}-\frac{2\pi n}{N}\right)\right]}\right]. \tag{21}$$

As in the case of N odd, we see that the values $x(2\pi l/N)$ obtained using zero-padding are exactly those that would be obtained from Eq. (2) for CST interpolation with N even and $\theta = 2\pi l/L$.

3. METHODS

3.1. Direct Fourier Image Reconstruction

The goal of two-dimensional tomographic image reconstruction is to estimate a two-dimensional distribution $f(x,y)$ of some object property, such as attenuation coefficient in computed tomography or the concentration of a radionuclide in emission tomography, from a set of one-dimensional projections of the distribution acquired at a number of different projection angles. Formally, the projection at projection angle φ is given by

$$P_\varphi(\xi) = \int_L f(\eta,\xi)d\eta , \tag{22}$$

where $f(\eta,\xi)$ represents the distribution in a coordinate system (η,ξ) rotated from the coordinate system (x,y) by an angle φ, and where L denotes the path of the line of integration. If we denote by $P_\varphi(u)$ the one-dimensional Fourier transform of $p_\varphi(\xi)$ with respect to ξ and by $F(\rho,\varphi)$ the two-dimensional Fourier transform of $f(x,y)$ in polar coordinates, the central slice theorem states that the two are related through

$$P_\varphi(\rho) = F(\rho,\varphi) \qquad \rho \geq 0, \quad 0 < \varphi < 2\pi. \tag{23}$$

Thus the FT of a projection at projection angle φ yields information about the 2D FT of the object along a spoke through the origin of Fourier space oriented at angle φ.

In practice, the projections are measured at a finite number N of angles $\varphi_n = n\pi/N$, $n = 0,...,N-1$, which we assume to be equally spaced over 180°, and each projection, assumed to be of length $2A$, is sampled at a finite number M of projections bins $\xi_m = 2mA/M, m = -(M-1)/2,...,(M-1)/2$. (For notational convenience, we take M to be odd throughout.) When transformed, these projections thus yield a finite set of samples of $F(\rho,\varphi)$, which we denote $F(\rho_m,\varphi_n)$, $\rho_m = m/2A$, $m = 0,...,(M-1)/2$ and $\varphi_n = n\pi/N$, $n = 0,...,2N-1$. Note that the number of angular samples on the polar grid is $2N$ and that the samples span 2π. This reflects the fact that a projection view at angle φ gives two samples of $F(\rho,\varphi)$, one at φ and one at $\varphi + \pi$. Specifically, the polar samples in the range $[0,\pi]$ correspond to the positive frequency

values of the transformed projections, while the polar samples in the range $[\pi, 2\pi]$ can be obtained from the negative frequency values of these transformed projections.

The true challenge in direct Fourier reconstruction is the interpolation from the polar grid above to a Cartesian grid $F(u_i, v_j)$. The size of the interpolated Cartesian grid should be chosen such that none of its points lie beyond the available polar samples; thus it should be the largest square that fits within a circle of radius $(M-1)/4A$. However, in order to ensure that the reconstructed image has the proper scale, this interpolated Cartesian grid should be padded with zeroes to length $(M-1)/A$ prior to performing the inverse 2D FFT. Having established the size of the Cartesian grid in frequency-space units, it must be subdivided into pixels in a way that eliminates or minimizes aliasing in the reconstructed image. To do so, the grid spacing must be no larger than $1/2A$ and preferably an integer divisor smaller.

3.2. Interpolation Between Coordinate Systems

The interpolation of the Cartesian grid values is generally accomplished by scanning the grid point by point, calculating the polar coordinate corresponding to each point and using known polar samples to interpolate a value at that polar point. This last step can be accomplished in many ways. Most crudely, the point could be simply be assigned the value of the nearest polar sample; this is known as nearest-neighbor interpolation. Slightly more accurately, bilinear interpolation could be performed using the four polar samples surrounding the point of interest.

Stark's approach is to perform CST interpolation in the angular direction and truncated Whittaker-Shannon sinc interpolation in the radial direction, obtaining an estimate $\hat{F}(\rho, \varphi)$ of $F(\rho, \varphi)$ at any point (ρ, φ) corresponding to a Cartesian grid point from the measured samples $F(\rho_m, \varphi_n)$ using

$$\hat{F}(\rho, \varphi) = \sum_{m=-(M-1)/2}^{(M-1)/2} \sum_{n=0}^{2N-1} F\left(\frac{m}{2A}, \frac{n\pi}{N}\right) \text{sinc}[2A(\rho - m/2A)] \frac{\sin\left[\frac{1}{2}(N-1)\left(\varphi - \frac{\pi n}{N}\right)\right]}{N \sin\left[\frac{1}{2}\left(\varphi - \frac{\pi n}{N}\right)\right]}. \tag{24}$$

The values of $F(m/2A, n\pi/N)$ for negative m can be determined from the known values for positive m using the relationship $F(-\rho, \varphi) = F^*(\rho, \varphi + \pi)$. This interpolation strategy is valid because the polar samples satisfy two crucial assumptions. First, because the projections are spatially compact, the radial functions on the polar grid should in principle be able to be reconstructed by infinite sinc interpolation from the samples with spacing $\Delta\rho = 1/2A$. The necessary truncation of the series is expected to introduce edge effects leading to high-frequency artifacts in the reconstructed image. Second, the angular functions, while clearly periodic with period 2π, are also expected to be sufficiently bandlimited that they can be sampled in accordance with the Nyquist criterion using a reasonable number of views [10], and can thus be exactly interpolated using the CST.

The reasons that validate Stark's strategy also justify the use of zero-padding interpolation to increase the density of polar samples prior to bilinear interpolation. Because the projections are spatially limited, they can be extended with zeroes to a factor of P times their original lengths prior to Fourier transforming. After transforming, the radial samples of the Fourier transform are P times as dense as previously. At this point, the angular samples at each fixed radial coordinate are subject to a one-dimensional Fourier transform. Because the functions are expected to be bandlimited in the angular direction, each resulting DFT sequence can be zero-padded as discussed in Sect. 2.2 to Q times its original length. Upon performing the inverse DFT, the angular samples of the FT are Q times as dense as previously. At this point interpolation proceeds as in the simple bilinear interpolation scheme discussed above: each Cartesian point is transformed to polar coordinates, and bilinear interpolation is performed among the four surrounding polar points.

3.3. Phantom studies

To compare these three Fourier-domain interpolation approaches, which we will refer to as the bilinear, zero-padding, and CST approaches, we first performed reconstructions of a Shepp-Logan head phantom (see Fig. 1a) in the absence of noise. The projection data consisted of 64 views spanning 180° and each containing 129 projection bins. We reconstructed a 129x129 pixel image (the choice of an odd number of bins and image pixels simplifies the implementation of the direct Fourier algorithms but is by no means necessary). In order to minimize aliasing artifacts, we interpolated onto a Cartesian grid of 257x257 pixels, having the same frequency-space extent as the transforms of the projections yet sampled twice as densely. The central 129x129 pixels of the inverse transform of this array then corresponded to the desired reconstructed image. When zero-padding, we applied zero-padding factors of 2 in both the angular and radial directions, thereby increasing the density of polar samples by a factor of 4.

For each reconstructed image we calculated the normalized root mean square distance of the reconstruction from the true phantom image. This accuracy measure is defined as

$$d = \left(\sum_{i=1}^{R} \sum_{j=1}^{R} (t_{ij} - r_{ij})^2 \Big/ \sum_{i=1}^{R} \sum_{j=1}^{R} (t_{ij} - \bar{t})^2 \right)^{1/2}, \qquad (25)$$

where t_{ij} and r_{ij} represent the pixel values of the pixel in the ith row and jth column of the true and reconstructed $R \times R$ images, respectively, and $\bar{t}$ denotes the average pixel value in the true image.

To examine the behavior of the algorithms in the face of noise we generated 500 sets of projections of a numerical torso phantom (see Fig. 1b) contaminated with Poisson noise (assuming 100,000 total counts in the sinogram). Each sinogram consisted of 64 projection angles of 65 projection bins spanning 180°. We reconstructed onto a 65x65 pixel array by inverting a Cartesian grid of 129x129 pixels having the same frequency-space extent as the transforms of the projections but sampled twice as densely. For each interpolation technique we calculated the mean of the 500 reconstructions, and the normalized RMS distance of this mean image from the true image. We also computed empirical variance images v_{ij} for each technique using

$$v_{ij} = \frac{1}{N} \left(\sum_{n=1}^{N} (r_{ij}^{(k)})^2 \right) - \left(\frac{1}{N} \sum_{n=1}^{N} r_{ij}^{(k)} \right)^2, \qquad (26)$$

where $r_{ij}^{(k)}$ is the value of the ijth pixel in the kth reconstruction and N is the total number of image realizations.

4. RESULTS AND DISCUSSION

The results of the noise-free reconstructions of the Shepp-Logan head phantom are shown in Fig. 2. Profiles of the row passing through the three small structures near the bottom of the image are also included. The normalized RMS distances and processing times for the three approaches are listed in Table 1. All three algorithms were implemented in Interactive Data Language, though the interpolation step of the circular sampling theorem interpolation was performed by a C subroutine. All computing was performed on an IBM RS6000-based workstation.

Table 1. Normalized RMS distances and processing times for reconstruction of the noise-free Shepp-Logan phantom.

Interpolation Method	Normalized RMS Distance	Processing Time
Bilinear	0.268	7.28 s
Zero-Padding	0.126	8.53 s
CST	0.148	2511.41 s

It is apparent, both qualitatively and quantitatively, that zero-padding interpolation produces images of comparable or superior quality to CST at a fraction of the computational cost. To be fair, CST can be made to run more rapidly by constraining the convolutions to smaller neighborhoods around the points of interest, but this comes at some cost in accuracy, and computing time still would barely approach that of the zero-padding approach, which benefits from the famed speed of the FFT. The excellent performance of the zero-padding approach relative to CST is explained in large part by their equivalent exactness in performing the angular interpolation and their equivalent approximations in performing the radial interpolation. So long as the polar samples satisfy periodicity and sampling criteria in the angular direction, the values interpolated in that direction by both methods are exact. In the radial direction, where the samples are not of a periodic function, the approximation entailed in using zero-padding interpolation is comparable to that involved in truncating the Whittaker-Shannon sinc interpolation series in Stark's CST approach. The error introduced in using bilinear interpolation in the final stage of the zero-padding approach can of course be made negligibly small by choosing sufficiently large zero-padding factors, though in practice, factors as small as 2 in each direction produce excellent results.

The accuracy of bilinear interpolation alone is, not surprisingly, worse than either of the other two methods. The reconstruction suffers from a depression in values toward the edges of the image, as can clearly be seen in the line profile. This rolloff phenomenon arises because of the deviation of the effective interpolation kernel from an ideal 2D-sinc interpolator—which the CST and zero-padding approaches approximate much more closely—and is discussed in greater detail by Jackson et al. [3].

The mean images of the noise-contaminated torso phantoms are illustrated in Fig. 3, along with line profiles passing through the bright annulus representing the heart near the top of the image. The normalized RMS distances listed in Table 2 demonstrate the same trends as in the noise-free reconstruction just discussed. Fig. 4 depicts the variance maps for the three approaches, along with profiles along the same line as illustrated in Fig. 3. The total variance in each map is listed in Table 2, where it can be seen that simple bilinear interpolation leads to a lower total empirical variance than the other two, whose

results are comparable. This is not surprising as the relatively crude bilinear interpolation smooths over noisy variations in Fourier space more than the CST and zero-padding approaches.

Table 2. Normalized RMS distances of mean images and total empirical variances for 500 reconstructions of a torso phantom containing Poisson noise (100,000 counts).

Interpolation Method	Normalized RMS Distance	Total Variance
Bilinear	0.252	2.78×10^6
Zero-Padding	0.122	4.16×10^6
CST	0.127	4.79×10^6

5. CONCLUSIONS

We have examined the use of three different methods for performing the interpolation from a polar to a Cartesian grid that is encountered in direct Fourier image reconstruction. Simple bilinear interpolation is, not surprisingly, found to yield poor-quality images, while a relatively efficient hybrid method, in which zero-padding interpolation is used to increase the density of polar samples prior to bilinear interpolation onto the Cartesian grid is found to yield comparable results to the much more computationally intense circular sampling theorem approach. We have attempted to account for the success of this hybrid approach in three ways. First, we have proved that zero-padding interpolation is exact and equivalent to circular sampling theorem interpolation when used to interpolate periodic functions sampled in accordance with the Nyquist criterion, which are precisely the sort that arise in the angular dimension of the polar grid. Second, we have pointed out that both methods make equivalent approximations in interpolating in the radial direction. Finally, we have indicated that the error arising in the subsequent use of bilinear interpolation in the hybrid approach can be minimized by increasing the density of polar samples sufficiently through zero-padding.

ACKNOWLEDGMENTS

REFERENCES

1.   G. T. Herman, *Image Reconstruction from Projections,* Academic Press, Orlando, 1980.
2.   J. D. O'Sullivan, "A fast sinc function gridding algorithm for Fourier inversion in computed tomography," *IEEE Trans. Med. Imaging* **4,** pp. 200-207, 1985.
3.   J. I. Jackson, C. H. Meyer, D. G. Nishimura, and A. Macovski, "Selection of a convolution function for Fourier inversion using gridding," *IEEE Trans. Med. Imaging* **10,** pp. 473-478, 1991.
4.   H. Stark, J. W. Woods, I. Paul, and R. Hingorani, "Direct Fourier reconstruction in computer tomography," *IEEE Trans. Acoust., Speech, Signal Processing* **29,** pp. 237-245, 1981.
5.   H. Stark, J. W. Woods, I. Paul, and R. Hingorani, "An investigation of computerized tomography by direct Fourier inversion and optimum interpolation," *IEEE Trans. Biomed. Eng.* **28,** pp. 496-505, 1981.
6.   S. X. Pan and A. C. Kak, "A computational study of reconstruction algorithms for diffraction tomography: interpolation versus filtered backpropagation," *IEEE Trans. Acoust., Speech, Signal Processing* **31,** pp. 1262-1275, 1983.
7.   H. Stark, "Sampling theorems in polar coordinates," *J. Opt. Soc. Am.* **69,** pp. 1519-1525, 1979.
8.   H. Stark and M. Wengrovitz, "Comments and corrections on the use of polar sampling theorems in CT," *IEEE Trans. Acoust., Speech, Signal Processing* **31,** pp. 1329-1331, 1983.
9.   D. Fraser, "Interpolation by the FFT revisited—an experimental investigation," *IEEE Trans. Acoust., Speech, Signal Processing* **37,** pp. 665-675, 1989.
10.   X. Pan, "Quasi band-limited properties of Radon transforms and their implications for increasing angular sampling densities," submitted to *IEEE Trans. Med. Imaging.*

Figure 1. Original numerical phantoms used in noise-free (a) and noise-contaminated (b) studies.
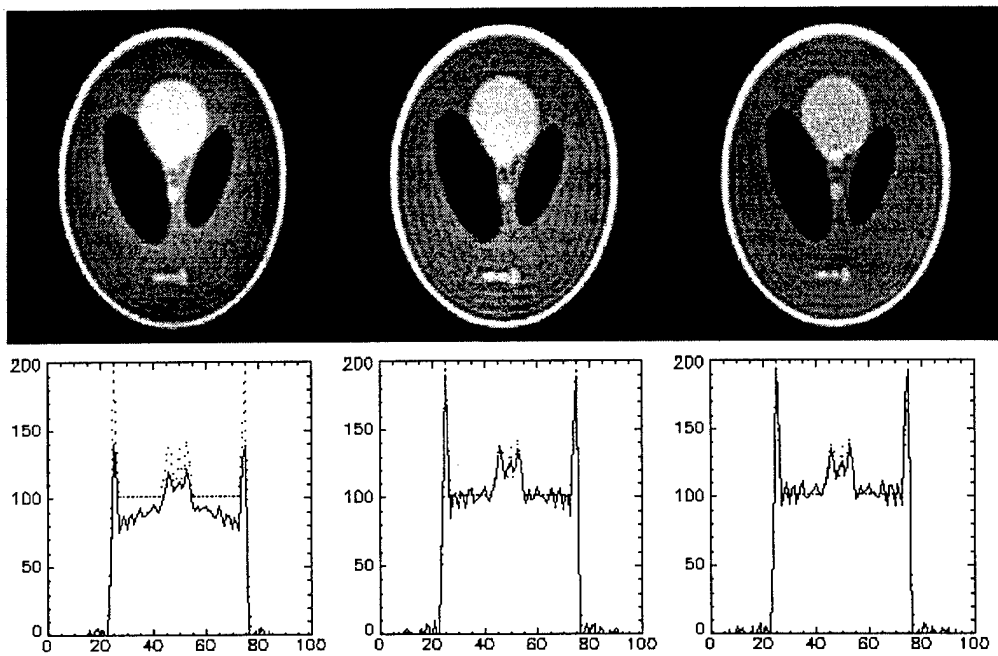

Figure 2. Direct Fourier reconstructions of the noise-free numerical Shepp-Logan phantom using simple bilinear interpolation (left), zero-padding interpolation followed by bilinear interpolation (center), with zero-padding factor 2 in both the angular and radial directions, and circular sampling theorem interpolation (right). The line profiles pass through the three small structures near the bottom of the images, with the dotted line representing the true values.
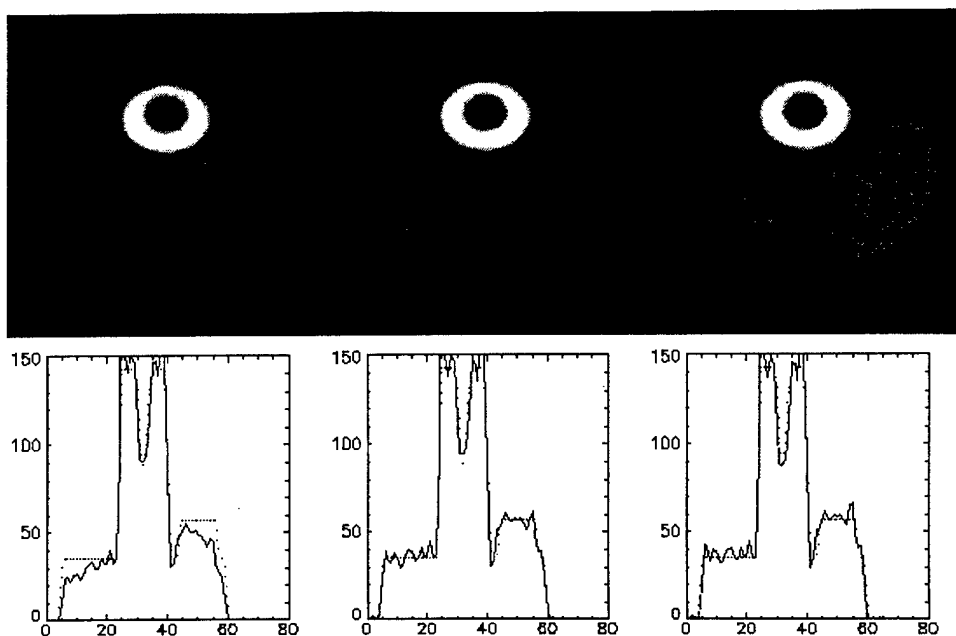
Figure 3. Means of 500 direct Fourier reconstructions of a numerical torso phantom contaminated with Poisson noise (100,000 counts) using bilinear interpolation (left), zero-padding interpolation followed by bilinear interpolation (center), with zero-padding factor 2 in both the angular and radial directions, and circular sampling theorem interpolation (right). The line profiles pass through the annulus near the top of the images, with the dotted line representing the true values.
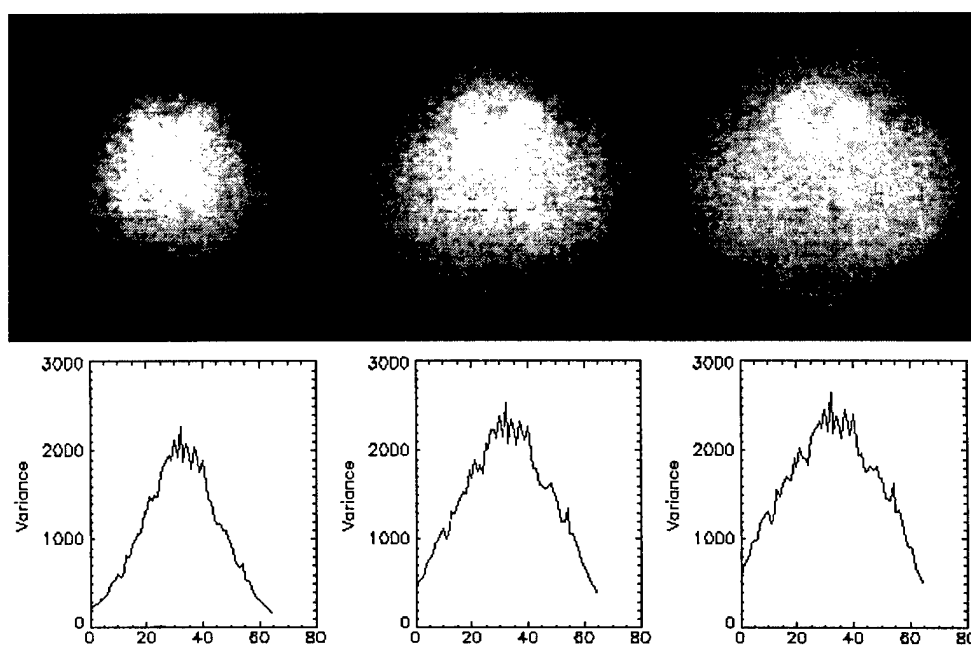
Figure 4. Variance images of 500 direct Fourier reconstructions of a numerical torso phantom contaminated with Poisson noise (100,000 counts) using bilinear interpolation (left), zero-padding interpolation followed by bilinear interpolation (center), with zero-padding factor =2 in both the angular and radial directions, and circular sampling theorem interpolation (right). The line profiles pass through the annulus near the top of the images.

Comparison of angular interpolation approaches in few-view tomography using statistical hypothesis testing

P. J. La Rivière and X. Pan

Department of Radiology, The University of Chicago, Chicago, IL

ABSTRACT

In this work we examine the accuracy of four periodic interpolation methods—circular sampling theorem interpolatiom, zero-padding interpolation, periodic spline interpolation, and linear interpolation with periodic boundary conditions—for the task of interpolating additional projections in a few-view sinogram. We generated 100 different realizations each of two types of numerical phantom—Shepp-Logan and breast—by randomly choosing the parameters that specify their constituent ellipses. Corresponding sinograms of 128 bins x 1024 angles were computed analytically and subsampled to 16, 32, 64, 128, 256, and 512 views. Each subsampled sinogram was interpolated to 1024 views by each of the methods under consideration and the normalized root-mean-square-error (NRMSE) with respect to the true 1024-view sinogram computed. In addition, images were reconstructed from the interpolated sinograms by FBP and the NRMSE with respect to the true phantom computed. The non-parametric signed rank test was then used to assess the statistical significance of the pairwise differences in mean NRMSE among the interpolation methods for the various conditions: phantom family (Shepp-logan or breast), number of measured views (16, 32, 64, 128, 256, or 512), and endpoint (sinogram or image). Periodic spline interpolation was found to be superior to the others in a statistically significant way for virtually every condition.

Keywords: Few-view tomography, interpolation, image reconstruction, sampling, zero-padding, spline, circular sampling theorem

1. INTRODUCTION

In nuclear-medicine tomographic studies, there is generally a strong correlation between imaging time and image quality. Longer acquisitions lead to better measurement statistics and thus to less noisy, more quantitatively accurate reconstructed images. These gains are negated, however, if the long acquisition times lead to patient motion and thus to motion artifacts in the reconstructed images. Such artifacts are especially common, or expected to be, in studies such as cardiac single-photon emission computed tomography (SPECT) and the anticipated dependent-breast SPECT scintimammography[1,2] where older, relatively inflexible patients need to assume an uncomfortable position. In these studies, the ability actually to *reduce* current imaging times without severely compromising basic image quality would likely result in better overall imaging performance due to the reduction of motion artifacts. As an added benefit, the reduced imaging times would increase patient throughput in busy clinics.

For a constant patient dose, imaging time can be reduced by reducing the number of angular projections acquired, by reducing the amount of time spent acquiring each projection, or by some combination of the two. While the distinction between the two is admittedly blurred when using a continuous acquisition mode, we have found in preliminary studies that so long as certain minimal angular sampling requirements are met, having fewer angular projections with more counts leads to a higher ideal-observer signal-to-noise ratio in the reconstructed images than does having more angular projections with fewer counts, given the same total number of counts. In the case of a step-and-shoot acquisition protocol, which is always used for gated cardiac SPECT, reducing the number of angular views has the additional benefit of reducing the amount of deadtime spent moving the camera between views. In this paper, then, we focus on algorithms tailored to generate diagnostically useful images from a smaller number of angular views than is usually used while holding the number of counts per view constant; that is, we focus on algorithms for *few-view* tomography.

The minimum number of angular views required to produce an accurate, artifact-free tomographic reconstruction of a given object is dictated by two factors. First, the angular sampling of the object's sinogram must satisfy, at least approximately, the Nyquist sampling condition.[3] Absent this, any reconstruction is doomed to suffer from angular aliasing artifacts. Second, the number of angular samples must satisfy the reconstruction algorithm's implicit assumptions about the density of angular sampling. For instance, it is well known that images reconstructed from a small number of angular views by filtered backprojection (FBP) are degraded by prominent star-shaped artifacts.

These two conditions are not in general equivalent, as can be appreciated most keenly when considering the case of imaging a circularly symmetric object. In this instance, a single projection view is sufficient to satisfy the Nyquist sampling condition, while a FBP reconstruction from this single view would be an uninterpretable set of parallel streaks. Indeed, Brooks et al.[4] have shown that for a circularly symmetric object imaged with n bins per projection, a minimum of $\sim 1.1\pi n/4$ projections must be acquired over 180° to produce an essentially artifact-free reconstruction using FBP.

The Nyquist condition is the more fundamental of the two sampling requirements discussed above, because when it is satisfied by a number of samples less than the number required by the reconstruction algorithm. it is in principle possible to interpolate exactly the additional views needed. This is only strictly true in the absence of noise, of course, a condition that rarely obtains in emission tomography. We have discussed elsewhere the use of a principled smoothing technique based on non-parametric regression with an explicit Poisson statistical model to control noise in each projection prior to the interpolation of additional angular views.[5] We have also examined the noise properties of various interpolation methods when confronted with noise-corrupted samples.[6] In the present work, then, we wish simply to address the question of which interpolation approach is most accurate for the interpolation problems encountered in few-view tomography in the absence of noise, or when noise has been controlled in a previous step.

Because the sinogram is periodic in the angular dimension, the interpolation method chosen should rightly be periodic. In Section 2.1 we review four periodic interpolation methods being evaluated: circular sampling theorem interpolation, zero-padding interpolation, periodic spline interpolation, and linear interpolation with periodic boundary conditions. In Section 2.2 we discuss the statistical hypothesis testing approach we have taken to judging the relative accuracy of the approaches in the face of various few-view tomography interpolation tasks. In Section 3 we present the results of these evaluations, and finally, in Section 4 we present our conclusions about which interpolation method is best suited to the task encountered in few-view tomography.

Finally, it should be mentioned that in this paper we focus exclusively on few-view reconstruction involving FBP even though iterative reconstruction algorithms, such as algebraic reconstruction techniques and maximum likelihood expectation-maximization, should in principle fare better when reconstructing from a small number of views because they make no implicit assumptions about the completeness or continuity of the projection dataset. While an exploration of the performance of iterative algorithms in the face of few-view datasets will be the subject of later work, we felt it important to begin with an examination of FBP, which remains the most widely used algorithm in clinical settings.

2. METHODS

2.1. Periodic Interpolation Methods

2.1.1. Circular sampling theorem and zero-padding interpolation

Circular sampling theorem (CST) interpolation is a special case of Whittaker-Shannon (W-S) sinc interpolation that applies to periodic functions.[7,8] Consider a periodic function $g(x)$ that has period X and which is bandlimited to frequency K (i.e., the coefficients of expansion a_k of the function's Fourier series satisfy $a_k = 0$ for $|k| > K$). Given $N \geq 2K + 1$ samples of $g(x)$ taken at points $x_n = nX/N$ ($n = 0, \ldots, N-1$) evenly spaced over one period, the CST states[8] that $g(x)$ can be interpolated exactly by use of

$$\widehat{g}(x) = \sum_{n=0}^{N-1} g(x_n)\sigma_N(x - x_n), \tag{1}$$

where

$$\sigma_N(x) = \sin[(2K + 1)\pi x/X]/N \sin(\pi x/X). \tag{2}$$

If the Nyquist condition is not satisfied, that is, if $g(x)$ is not truly bandlimited to frequency K or if $N < 2K+1$, Eqs. (1) and (2) no longer represent exact interpolation, but they remain mathematically meaningful and, often, practically useful. For instance, if the spectral components beyond frequency K are negligibly small but not exactly zero, interpolating with Eqs. (1) and (2) remains very accurate.

A second periodic interpolation approach, zero-padding (ZP) interpolation, involves extending the discrete Fourier transform (DFT) of a finite sequence with zeroes and taking an inverse DFT to generate a more densely sampled

version of the original sequence with values interpolated at intermediate positions between the original measured samples.[9-12] Specifically, one begins by taking the DFT of the sequence $g(x_n)$, which is given by

$$c_k = \frac{1}{N} \sum_{n=0}^{N-1} g(x_n) \exp(-j2\pi nk/N), \tag{3}$$

for $k = 0, \ldots, N-1$, where $j = \sqrt{-1}$. Zero-padding involves the creation of a new sequence $d_{k'}$, having $L = PN$ elements (where P is an integer). If $g(x)$ is assumed to be bandlimited to frequency K and if $N \geq 2K + 1$, the sequence $d_{k'}$ is defined as follows:

$$d_{k'} = \begin{cases} c_{k'} & k' = 0, \ldots, K \\ 0 & k' = K+1, \ldots, L-K-1 \\ c_{k'-L+N} & k' = L-K, \ldots, L-1. \end{cases} \tag{4}$$

A more densely sampled sequence $\widehat{g}(x_l)$, where $x_l = lX/L$ ($l = 0, \ldots, L-1$), is now obtained by taking the inverse DFT of the sequence $d_{k'}$,

$$\widehat{g}(x_l) = \sum_{k'=0}^{L-1} d_{k'} \exp(j2\pi lk'/L), \tag{5}$$

for $l = 0, \ldots, L-1$. ZP interpolation is generally viewed as a somewhat crude interpolation approach, but this reputation is undeserved. It can be shown that ZP interpolation is equivalent to CST interpolation, in that the spatial-domain interpolation function corresponding to the ZP operation in frequency space is just $\sigma_N(x)$ given in Eq. (2).[13] That is, it can be shown that

$$\widehat{g}(x_l) = \sum_{n=0}^{N-1} g(x_n) \sigma_N(x_l - x_n), \tag{6}$$

where $\sigma_N(x)$ is given by Eq. (2). This equivalence holds regardless of whether the Nyquist condition is satisfied. If it is satisfied, then ZP interpolation, like CST interpolation, is exact. Obviously, the CST interpolation formula can be used to estimate $g(x)$ at any arbitrary point x whereas zero-padding interpolation is constrained to interpolate onto a fixed, equispaced grid P times denser than the original samples. However, since ZP yields a continuous interpolated curve in the limit as $P \to \infty$, we shall treat these two approaches as one in the subsequent analysis, using the continuous form of Eq. (1) to represent them both. Given this equivalence, ZP is in general to be preferred in practice because it is considerably more computationally efficient than the CST. Whenever using ZP/CST below, we take K to be $N/2 - 1$; that is, we assume that the bandlimit of the function is equal to (or higher than) than the Nyquist frequency of the samples.

2.1.2. Periodic spline interpolation

Splines are piecewise cubic polynomials that are continuous up to and including the second derivative at the joints between pieces.[14,15] Periodic splines are further constrained to satisfy periodic boundary conditions. A spline $\widehat{g}(x)$ can be represented by

$$\widehat{g}(x) = a_n + b_n(x - x_n) + c_n(x - x_n)^2 + d_n(x - x_n)^3, \tag{7}$$

for $x \in [x_n, x_{n+1}]$, where the x_n are the abscissas at which the data is measured and $n = 0, \ldots, N-1$. Fitting a spline is thus tantamount to finding the coefficients a_n, b_n, c_n, and d_n subject to interpolation, continuity, and boundary conditions. A number of efficient algorithms exist for doing so.[14]

2.1.3. Linear interpolation with periodic boundary conditions

The last approach to be considered is also the simplest: linear interpolation with periodic boundary conditions. In this case the interpolation takes the form

$$\widehat{g}(x) = (1 - w_n(x))g(x_n) + w_n(x)g(x_{n+1}), \tag{8}$$

for $x \in [x_n, x_{n+1}]$, where the x_n are the abscissas at which the data is measured, $n = 0, \ldots, N-1$, and $w_n(x) = (x - x_n)/(x_{n+1} - x_n)$. The peridocity condition enters when $n = N-1$, in which case $g(x_N)$ is taken to be equal to $g(x_o)$.
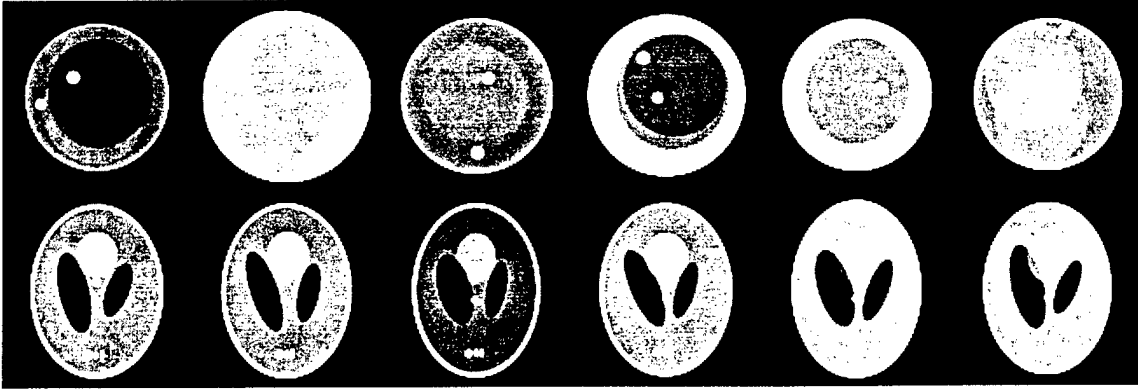
Figure 1. Six typical realizations of the breast (upper) and Shepp-Logan (lower) types of phantoms whose sinograms were used in the interpolation accuracy studies. The realizations were generated by choosing the parameters governing the phantoms' constituent ellipses at random according to predetermined probability laws.

2.2. Evaluation of Accuracy Using Statistical Hypothesis Testing

Various theoretical claims can be made about the accuracy of the interpolation methods being considered, each depending on the degree to which the data satisfy the assumptions underlying the method. For instance, linear interpolation is exact if the true function is piecewise linear between the measured samples and quite accurate if the function is sampled densely enough that it is nearly linear between measured samples. The same holds for spline interpolation and piecewise cubic. or nearly so, functions. CST and ZP interpolations, as noted above, are exact for periodic functions that are bandlimited to the Nyquist frequency of the samples. Another approach to the evaluation of interpolation methods involves examining the Fourier transforms of the methods' interpolation kernels to see how well they approximate the rectangular transform of the theoretically exact sinc kernel. This again implicitly assumes that the function is bandlimited to the Nyquist frequency of the samples, or, at least, that violations of the bandlimited assumption only negligibly compromise the accuracy of the approach.

Real data rarely satisfy any of these assumptions exactly. and in practice we have found that mild violations, particularly of the bandlimited assumption. can lead to undesirable errors. We wished to judge the methods being considered, then, not on the basis of theoretical claims but empirically in the context of the task of interest: the interpolation of additional angular views in a sinogram. To do so, we required numerical phantoms whose projections could be calculated analytically. In this way, we could generate a sinogram with a reduced number of views, interpolate to a larger number of views, and then compare quantitatively the interpolated sinogram to the exact sinogram for the corresponding number of views. Fortunately, it is possible to compute the projections an ellipse of arbitrary size, position, and orientation exactly.[16] allowing much freedom in the design of objects whose sinograms were to be interpolated. We focused on two types of phantoms: the familiar Shepp-Logan brain phantom[17] and a simulated breast phantom consisting of a large outer circle. some slightly smaller background ellipses, and small circular lesions.

Simply comparing the success of the approaches in interpolating a single sinogram each for one or two canonical phantoms would provide more anecdotal than genuinely rigorous evidence on which to base the choice of interpolation method for few-view tomography. The outcome could depend as much on numerical happenstance as on the genuine strengths of the approaches. Instead, we generated 100 different "realizations" of each of the two types of numerical phantom by choosing the parameters specifying the constituent ellipses of each type to vary according to predetermined probability laws. Some typical phantom realizations are illustrated in Figure 1.

For each realization of the phantom parameters we generated a parallel-beam sinogram of 128 bins and 1024 angular views. We then subsampled the sinogram to generate sinograms of 16, 32, 64, 128, 256, and 512 views. Each of these subsampled sinograms was interpolated to 1024 views by applying each of the three periodic interpolation approaches discussed in turn. The resulting interpolated sinograms were then compared to the analytically computed 1024-view sinogram and the normalized root-mean-square error (NRMSE) between the two computed. This error measure is defined as

$$d = \left(\sum_{i=1}^{R} \sum_{i=1}^{R} (t_{ij} - r_{ij})^2 / \sum_{i=1}^{R} \sum_{i=1}^{R} (t_{ij} - \bar{t})^2 \right)^{1/2}, \tag{9}$$

	ZP vs. Spline			Spline vs. Linear			Linear vs. ZP		
# views	Mean Diff.	W_+	p-value	Mean Diff.	W_+	p-value	Mean Diff.	W_+	p-value
16	1.09×10^{-3}	29.0	0.0000	4.96×10^{-4}	1234.0	0.0000	-1.58×10^{-3}	4572.0	0.0000
32	6.98×10^{-4}	12.0	0.0000	-4.40×10^{-4}	4286.0	0.0000	-2.58×10^{-4}	2946.0	0.0739
64	3.10×10^{-4}	0.0	0.0000	-3.91×10^{-4}	5024.0	0.0000	8.19×10^{-5}	1882.0	0.0135
128	1.72×10^{-4}	4.0	0.0000	-1.51×10^{-4}	5003.0	0.0000	-2.0×10^{-5}	3040.0	0.0383
256	9.06×10^{-5}	0.0	0.0000	-6.86×10^{-5}	4890.0	0.0000	-2.2×10^{-5}	3664.0	0.0000
512	4.83×10^{-5}	1.0	0.0000	-2.95×10^{-5}	4835.0	0.0000	-1.88×10^{-5}	4239.0	0.0000

Table 1. Pairwise mean breast phantom sinogram NRMSE differences, test statistics, and p-values for the three interpolation methods.

where t_{ij} and r_{ij} represent the pixel values of the pixel in the ith row and jth column of the true and reconstructed $R \times R$ images, respectively, and $\bar{t}$ denotes the average pixel value in the true image. In order to see how the sinogram interpolation error affected the accuracy of reconstructed images, we also used FBP to reconstruct an image from each sinogram interpolated by each of the three approaches, and then calculated the NRMSE with respect to the true phantom.

The result of this simulation was a set of 100 NRMSE values for each combination of phantom family (Shepp-logan or breast), interpolation method (CST/ZP, spline, or linear), number of measured views (16, 32, 64, 128, 256, or 512), and endpoint (sinogram or image). This data lent itself naturally to a statistical hypothesis testing analysis seeking to evaluate for each task (phantom family, number of measured angles, and endpoint) the null hypothesis that, considered pairwise, the three interpolation methods yielded the same NRMSE. Refuting the null hypothesis would allow the methods to be ranked for each task.

Because the interpolation methods were applied in parallel to the same realizations of the phantom, a paired test must be used to evaluate the null hypotheses. The familiar paired t-test, in which the differences between paired samples are compared to cutoffs of the t-distribution, implicitly assumes that the differences are a sample from a normal distribution. As exploratory data analysis indicated that the distribution of differences was not evidently normal in the present case, we turned instead to the nonparametric signed rank (SR) test for paired samples.[18] In the SR test, the absolute values of the differences are ranked (with appropriate handling of ties), the signs of the differences applied to the corresponding ranks, and the sum W_+ of those ranks having positive signs computed. The distribution of W_+ under the null hypothesis that the two conditions produce the same values can be easily computed, and extreme values of W_+ signal that one condition tends to produce larger value than the other. Finally, the p-value (the probability that a W_+ as extreme or more extreme than the one found could arise under the null hypothesis) of each W_+ can be computed from this distribution as well.

3. RESULTS

The results for the interpolation of the family of breast phantoms are depicted in Figure 2. For each interpolation method, Figure 2(a) plots the mean sinogram NRMSE for the 100 realizations versus the number of measured projections. A higher NRMSE corresponds to worse accuracy. Because the curves in Figure 2(a) are difficult to distinguish, particularly for large numbers of measured projections, the NRMSE relative to the periodic spline NRMSE at each number of measured projections is plotted in Figure 2(b). The corresponding curves for reconstructed image NRMSE are depicted in Figures 2(c) and 2(d). Tables 1 and 2 list the mean difference between each pair of methods for each number of measured projections as well as the corresponding value of the SR statistic W_+ and its two-sided p-value. A low p-value indicates that the mean difference between the two methods is statistically significant, and we adopt throughout a stringent 0.01 cutoff for statistical significance.

A few trends common to all three interpolation methods can be gleaned from Figure 2. For one, the accuracy of all three approaches, as measured by the sinogram NRMSE, improves rapidly as the number of measured projections increases. This is not surprising—the assumptions underlying all three approaches are better satisfied as the number of samples increases—but it does offer an interesting contrast to the trend seen in reconstructed image NRMSE. These NRMSE values level off beyond 64 samples, with the change in overall NRMSE from one number of measured
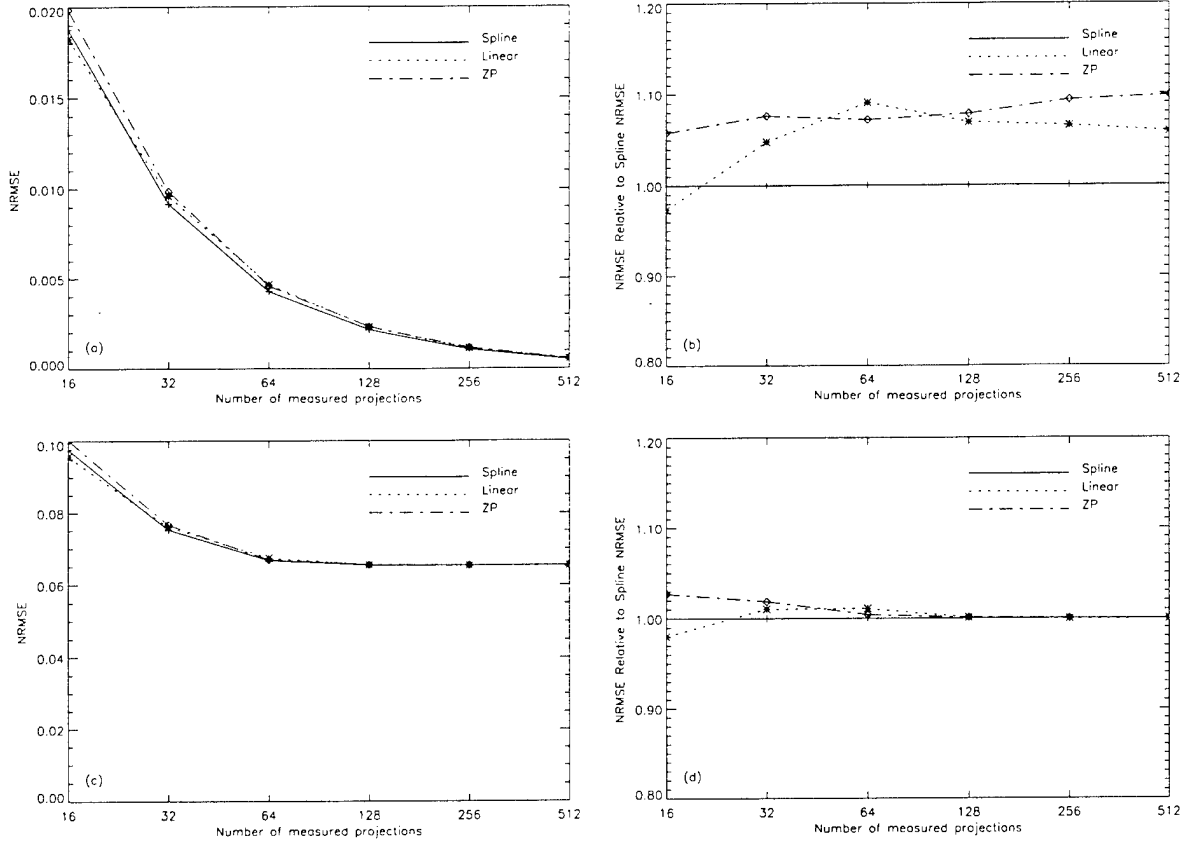
Figure 2. NRMSE and relative NRMSE plots for the family of breast phantoms. (a) Plot of sinogram NRMSE vs. number of measured projections for each interpolation method. (b) Plot of sinogram NRMSE relative to spline sinogram NRMSE vs. number of measured projections for each interpolation method. (c) Plot of reconstructed image NRMSE vs. number of measured projections for each interpolation method. (d) Plot of reconstructed image NRMSE relative to spline image NRMSE vs. number of measured projections for each interpolation method.

	ZP vs. Spline			Spline vs. Linear			Linear vs. ZP		
# views	Mean Diff.	W_+	p-value	Mean Diff.	W_+	p-value	Mean Diff.	W_+	p-value
16	2.65×10^{-3}	88.0	0.0000	1.91×10^{-3}	966.0	0.0000	-4.56×10^{-3}	4520.0	0.0000
32	1.40×10^{-3}	147.0	0.0000	-7.76×10^{-4}	3771.0	0.0000	-6.26×10^{-4}	2659.0	0.3225
64	2.54×10^{-4}	138.0	0.0000	-7.03×10^{-4}	4237.0	0.0000	4.49×10^{-4}	1444.0	0.0001
128	6.74×10^{-5}	3.0	0.0000	-5.33×10^{-5}	2281.0	0.2008	-1.41×10^{-5}	3227.0	0.0079
256	3.30×10^{-6}	75.0	0.0000	2.69×10^{-5}	1005.0	0.0000	-3.02×10^{-5}	4138.0	0.0000
512	2.00×10^{-7}	1289.5	0.0000	7.90×10^{-6}	509.0	0.0000	-8.00×10^{-6}	4466.5	0.0000

Table 2. Pairwise mean breast phantom image NRMSE differences, test statistics, and p-values for the three interpolation methods.

projections to the next being comparable to the differences among the interpolation method for a given number of measured projections. This leveling off is of course due to the fact that errors associated with other inacurracies in the image reconstruction process, such as finite radial sampling, come to overwhelm the error due to finite angular sampling as the number of angular samples grows large.

As for the relative performance of the three interpolation approaches, when comparing sinogram NRMSEs, periodic spline interpolation is seen to be superior to the other two in a statistically significant way for all but the
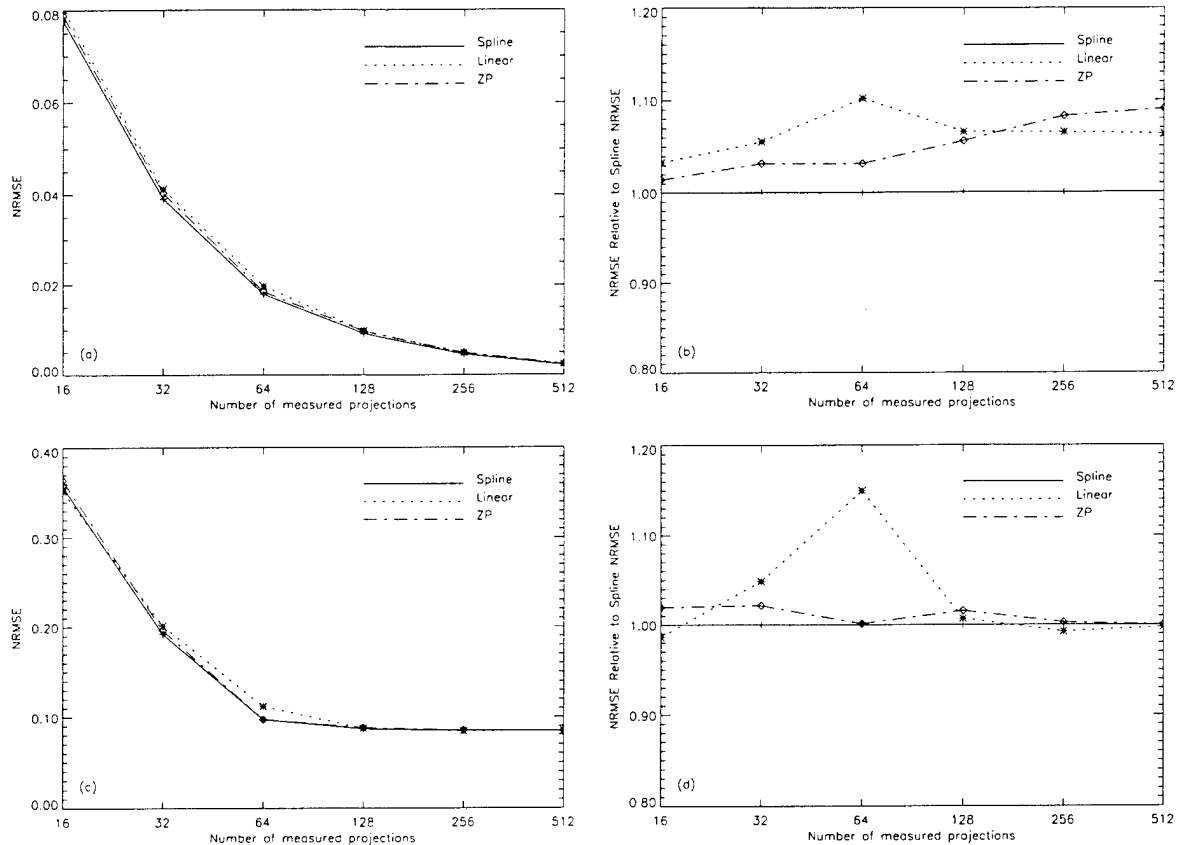
Figure 3. NRMSE and relative NRMSE plots for the family of Shepp-Logan phantoms. (a) Plot of sinogram NRMSE vs. number of measured projections for each interpolation method. (b) Plot of sinogram NRMSE relative to spline sinogram NRMSE vs. number of measured projections for each interpolation method. (c) Plot of reconstructed image NRMSE vs. number of measured projections for each interpolation method. (d) Plot of reconstructed image NRMSE relative to spline image NRMSE vs. number of measured projections for each interpolation method.

smallest number of samples—16—where linear interpolation performs best. Linear interpolation is seen to be better than or statistically indistinguishable from ZP interpolation for all numbers of measured projections. Similar trends are observed in comparing image NRMSEs. Spline outperforms ZP interpolation in every case, and is better than linear interpolation for moderate numbers of measured projections, but slightly worse or indistiguishable for very large numbers of samples. Linear is again seen to be better than or indistinguishable from ZP for all numbers of measured projections except 64. It may at first seem paradoxical that comparing image NRMSEs does not always yield the same relative performance as sinogram NRMSEs. Shouldn't more accurate sinograms necessarily produce more accurate images? The resolution lies in the fact that NRMSEs are global error measures obtained by summing over local errors. It is the local errors that propagate directly into the reconstructed images and these may combine or cancel differently depending on their relative positions in the sinogram. Thus the geography of the local sinogram errors, as well as their sum, influences the global error in the reconstruction, and so a sinogram with a higher NRMSE than another can in principle lead to a reconstructed image with a lower NRMSE than the other.

The corresponding results for the family of Shepp-Logan phantoms are illustrated in Figure 3 and Tables 3 and 4. The same trends are observed with regard to the overall performance of the three approaches. The sinogram NRMSEs fall rapidly with increasing number of measured projections, while the reconstructed image NRMSEs level off rapidly beyond 64 projections. While these trends are as before, it should be pointed out that the magnitude of the NRMSEs for both sinograms and reconstructed images is significantly higher for the Shepp-Logan type phantoms than it was for the breast phantoms, owing to the fact that the Shepp-Logan phantoms are more complex and their sinograms more variable in the angular direction than the breast phantoms.

# views	ZP vs. Spline			Spline vs. Linear			Linear vs. ZP		
	Mean Diff.	W_+	p-value	Mean Diff.	W_+	p-value	Mean Diff.	W_+	p-value
16	1.12×10^{-3}	0.0	0.0000	-2.58×10^{-3}	5045.0	0.0000	1.45×10^{-3}	575.0	0.0000
32	1.24×10^{-3}	53.0	0.0000	-2.17×10^{-3}	5032.0	0.0000	9.18×10^{-4}	1530.0	0.0003
64	5.72×10^{-4}	747.0	0.0000	-1.83×10^{-3}	4503.0	0.0000	1.26×10^{-3}	1252.0	0.0000
128	5.27×10^{-4}	9.0	0.0000	-6.19×10^{-4}	4945.0	0.0000	9.20×10^{-5}	2136.0	0.0905
256	3.82×10^{-4}	1.0	0.0000	-3.06×10^{-4}	4845.0	0.0000	-7.58×10^{-5}	3102.0	0.0236
512	1.94×10^{-4}	0.0	0.0000	-1.39×10^{-4}	4722.0	0.0000	-5.52×10^{-5}	3203.0	0.0099

Table 3. Pairwise mean Shepp-Logan phantom sinogram NRMSE differences, test statistics. and p-values for the three interpolation methods.

# views	ZP vs. Spline			Spline vs. Linear			Linear vs. ZP		
	Mean Diff.	W_+	p-value	Mean Diff.	W_+	p-value	Mean Diff.	W_+	p-value
16	6.91×10^{-3}	0.0	0.0000	4.88×10^{-3}	244.0	0.0000	-1.18×10^{-2}	5050.0	0.0000
32	4.10×10^{-3}	177.0	0.0000	-9.31×10^{-3}	5050.0	0.0000	5.21×10^{-3}	968.0	0.0003
64	1.07×10^{-4}	2454.0	0.4036	-1.46×10^{-2}	5050.0	0.0000	1.44×10^{-2}	0.0	0.0000
128	1.36×10^{-3}	0.0	0.0000	-5.95×10^{-4}	3909.0	0.0000	-7.68×10^{-4}	4153.0	0.0905
256	2.19×10^{-4}	0 .0	0.0000	6.57×10^{-4}	0.0	0.0000	-8.77×10^{-4}	5050.0	0.0236
512	9.00×10^{-6}	1.0	0.0000	2.12×10^{-4}	0.0	0.0000	-2.21×10^{-4}	5050.0	0.0099

Table 4. Pairwise mean Shepp-Logan phantom image NRMSE differences, test statistics, and p-values for the three interpolation methods.

As for the relative performance of the interpolation methods for the Shepp-Logan phantoms, the results are similar but not identical to those of the breast phantom. When comparing sinogram NRMSEs, periodic spline interpolation is seen to be better than ZP or linear interpolation in a statistically significant way for all numbers of measured projections. Linear interpolation is better than ZP for larger numbers of projections (256 or 512), worse for small numbers (16, 32, and 64). and indistingushable for 128 projections. When comparing reconstructed image NRMSEs, spline is still better than ZP for every number of measured projections, and, as before, better than linear interpolation for moderate numbers of projections (32, 64, and 128) while worse for extreme numbers (16, 256, and 512). Linear is better than zero-padding for extreme numbers of measured projections (16, 128, 256, and 512) and worse for moderate numbers (32 and 64).

4. DISCUSSION AND CONCLUSIONS

We have presented a study of the empirical accuracy of three different interpolation approaches for the task of interpolating additional angular views in a sinogram. For two types of numerical phantom—Shepp-Logan and breast—we generated 100 different realizations by randomly choosing the parameters that specify their constituent ellipses. Sinograms of 128 bins x 1024 angles were computed analytically and subsampled to 16, 32, 64, 128, 256, and 512 views. Each subsampled sinogram was interpolated to 1024 views by each of the three methods under consideration—CST/ZP, periodic spline, and linear interpolation—and the NRMSE compared to the true 1024-view sinogram computed. The non-parametric signed rank test was then used to assess the statistical significance of the pairwise differences in mean NRMSE among the interpolation methods for the various conditions: phantom family (Shepp-logan or breast), number of measured views (16, 32, 64, 128, 256, or 512), and endpoint (sinogram or image).

Periodic spline interpolation was found to be superior in a statistically significant way to CST/ZP and linear interpolation for virtually every condition. It was superior to CST/ZP in every instance, and fell short of linear interpolation only in the case of a very small number of measured angles (16) for breast sinogram NRMSE and for extremely large or small numbers of angles for reconstructed image NRMSE for both types of phantoms. It was certainly always superior for interpolation from the moderate number of angles—32 and 64—likely to be relevant to few-view tomography. The strong performance of spline interpolation can be attributed to the combination of its

great flexibility and relatively local response. That is, while splines can capture subtle features that frequently elude piecewise linear curves, they do not allow sharp edges that they are unable to capture faithfully to influence unduly the accuracy of the curve at distant points. CST/ZP interpolation, on the other hand, while blessed with considerably more flexibility than linear interpolation, tends to suffer from widespread ringing artifacts (Gibbs phenomena)[19] when sharp edges cause a violation, however mild, of the bandlimited assumption on which it is built.

Indeed, the relatively poor performance of CST/ZP interpolation was one of the surprises in this work, and careful examination of individual curves interpolated by this method indicated that Gibbs artifacts were indeed the principal cause of the poor performance. Ironically, while the severity of the ringing increases as the number of measured samples decreases, zero-padding performed relatively well in this range compared to linear interpolation (though not to spline), and generally fell behind linear only for denser sampling. This is likely because at this low sampling density the piecewise linear approximation is so poor over most of the function being interpolated that CST/ZP, with its more flexible curves, performs better despite the ringing. The ringing only becomes a factor as both methods perform similarly over the slowly varying parts of the curve, with linear interpolation's local response around edges then being preferable to the ringing introduced by CST/ZP. Attempts to mitigate the ringing using anti-oscillation filters such as Lanczos's sigma filters[20] met with some success, but the smoothing caused by these filters degraded the NRMSE performance at least as much as the ringing had. In short, unless the few-view data is known to be precisely bandlimited, CST/ZP interpolation is not recommended.

This phenomenon of non-local response may also explain linear interpolation's suprising success relative to spline interpolation for large numbers (256 or 512) of measured projections. With sampling of such density, the piecewise linear approximation is of course very good over most of the curve, and those few points where the approximation is not very good contribute little to the total NRMSE. The accuracy of spline interpolation, on the other hand, is likely to be comparable to linear interpolation over most of the interpolated points, but may suffer more widely the effects of an edge because the spline interpolation kernel is considerably wider than that of linear interpolation. Nonetheless, spline interpolation was seen to be clearly superior for both simple and complex objects in precisely the angular sampling range (32-64) of interest to few-view tomography.

ACKNOWLEDGMENTS

REFERENCES

1. P. J. La Rivière, X. Pan, and B. C. Penney, "Ideal-observer analysis of lesion detectability in planar, conventional SPECT, and dedicated SPECT scintimammography using effective multi-dimensional smoothing," *IEEE Trans. Nucl. Sci.* **45**, pp. 1273-1279, 1998.

2. H. Wang, C. Scarfone, K. L. Greer, R. E. Coleman, and R. J. Jaszczak, "Prone breast tumor imaging using vertical axis-of-rotation SPECT systems: An initial study," *IEEE Trans. Nucl. Sci.* **44**, pp. 1271-1276, 1997.

3. R. J. Marks, *Introduction to Shannon Sampling and Interpolation Theory*, Springer Verlag, New York, 1991.

4. R. A. Brooks, G. H. Weiss, and A. J. Talbert, "A new approach to interpolation in computed tomography," *J. Comput. Assist. Tomogr.* **2**, pp. 577-585, 1978.

5. P. J. La Rivière and X. Pan, "Few-view tomography using interpolating and smoothing splines with implications for cardiac spect," in *Proc. IEEE Nucl. Sci. Symp. and Med. Imag. Conf.*, 1999. (To appear).

6. P. J. La Rivière and X. Pan, "Noise properties of periodic interpolation methods with implications for few-view tomography," in *Proc. IEEE Nucl. Sci. Symp. and Med. Imag. Conf.*, 1999. (To appear).

7. H. Stark, "Sampling theorems in polar coordinates," *J. Opt. Soc. Am.* **69**, pp. 1519-1525, 1979.

8. H. Stark and M. Wengrovitz, "Comments and corrections on the use of polar sampling theorems in CT," *IEEE Trans. Acoust., Speech, Signal Processing* **31**, pp. 1329-1331, 1983.

9. R. W. Schafer and L. R. Rabiner, "A digital signal processing approach to interpolation," *Proc. IEEE* **61**, pp. 692-702, 1973.

10. K. P. Prasad and P. Satyanarayana, "Fast interpolation algorithm using FFT," *Electron. Lett.* **22**, pp. 185-187, 1986.

11. D. Fraser, "Interpolation by the FFT revisited—an experimental investigation," *IEEE Trans. Acoust., Speech. Signal Processing* **37**, pp. 665–675, 1989.

12. T. J. Cavicchi, "DFT time-domain interpolation," *IEE Proceedings-F* **139**, pp. 207–211, 1992.

13. P. J. La Rivière and X. Pan, "Mathematical equivalence of zero-padding and circular sampling thorem interpolation with implications for direct Fourier image reconstruction," in *Proc. SPIE*, vol. 3338, pp. 1117–1126, 1998.

14. H. Spath, *One-Dimensional Spline Interpolation Algorithms*, A. K. Peters, Wellesley, MA, 1993.

15. C. H. Reinsch, "Smoothing by spline functions," *Numer. Math.* **10**, pp. 177–183, 1967.

16. A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Academic Press, New York, 1982.

17. L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," *IEEE Trans. Nucl. Sci.* **21**, pp. 21–43, 1974.

18. J. A. Rice, *Mathematical Statistics and Data Analysis*, Wadsworth, Belmont, CA, 1994.

19. E. O. Brigham, *The Fast Fourier Transform and Its Applications*, Prentice Hall, Englewood Cliffs, 1988.

20. C. Lanczos, *Applied Analysis*, Dover, Mineola. NY, 1988.

Noise Properties of Periodic Interpolation Methods with Implications for Few-View Tomography

P. J. La Rivière, *Student Member, IEEE,* and X. Pan, *Member, IEEE*

Department of Radiology, The University of Chicago, Chicago, IL, 60637

Abstract

A number of methods exist specifically for the interpolation of periodic functions from a finite number of samples. When the samples are known exactly, exact interpolation is possible under certain conditions, such as when the function is bandlimited to the Nyquist frequency of the samples. However, when the samples are corrupted by noise, it is just as important to consider the noise properties of the resulting interpolated curve as it is to consider its accuracy. In this work, we derive analytic expressions for the covariance and variance of curves interpolated by three periodic interpolation methods—circular sampling theorem, zero-padding, and periodic spline interpolation—when the samples are corrupted by noise. We perform empirical studies for the special cases of white and Poisson noise and find the results to be in agreement with the analytic derivations. The implications of these findings for few-view tomography are also discussed.

I. INTRODUCTION

The need to interpolate samples of periodic functions arises in a number of important medical imaging applications. For instance, when performing emission computed tomography imaging of a compact, reasonably symmetric object, such as the breast, one can achieve adequate angular sampling of the object's sinogram with a relatively small number of projection views. However, using filtered backprojection (FBP) to reconstruct the image may still lead to star-shaped artifacts, because FBP implicitly requires a relatively high density of angular samples [1]. In these situations, periodic interpolation may be used to interpolate additional angular views between the measured ones in order to satisfy FBP's sampling requirements. The need to perform periodic interpolation also arises in direct Fourier image reconstruction, where it is necessary to interpolate from a polar to a Cartesian grid in Fourier space [2]. The interpolation is often accomplished using separate 1D interpolations in the radial and azimuthal directions, and the azimuthal interpolation should rightly be periodic.

There exist a number of methods for interpolating periodic functions. Circular sampling theorem (CST) interpolation, for one, is a special case of Whittaker-Shannon (W-S) sinc interpolation that applies to periodic functions [3, 4]. Consider a periodic function $g(x)$ that has period X and which is bandlimited to frequency K (i.e., the coefficients of expansion a_k of the function's Fourier series satisfy $a_k = 0$ for $|k| > K$). Given $N \geq 2K + 1$ samples of $g(x)$ taken at points $x_n = nX/N$ ($n = 0, \ldots, N - 1$) evenly spaced over one period, the CST states [4] that $g(x)$ can be interpolated exactly

by use of

$$g(x) = \sum_{n=0}^{N-1} g(x_n)\sigma_N(x - x_n), \tag{1}$$

where

$$\sigma_N(x) = \sin[(2K + 1)\pi x/X]/N \sin(\pi x/X). \tag{2}$$

If the Nyquist condition is not satisfied, that is, if $g(x)$ is not truly bandlimited to frequency K or if $N < 2K + 1$, Eqs. 1 and 2 no longer represent exact interpolation, but they remain mathematically meaningful and, often, practically useful. For instance, if the spectral components beyond frequency K are negligibly small but not exactly zero, interpolating with Eqs. 1 and 2 remains very accurate.

A second periodic interpolation approach, zero-padding (ZP) interpolation, involves extending the discrete Fourier transform (DFT) of a finite sequence with zeroes and taking an inverse DFT to generate a more densely sampled version of the original sequence with values interpolated at intermediate positions between the original measured samples [5–8]. Specifically, one begins by taking the DFT of the sequence $g(x_n)$, which is given by

$$c_k = \frac{1}{N} \sum_{n=0}^{N-1} g(x_n) \exp(-j2\pi nk/N), \tag{3}$$

for $k = 0, \ldots, N - 1$, where $j = \sqrt{-1}$. Zero-padding involves the creation of a new sequence $d_{k'}$, having $L = PN$ elements (where P is an integer). If $g(x)$ is assumed to be bandlimited to frequency K and if $N \geq 2K + 1$, the sequence $d_{k'}$ is defined as follows:

$$d_{k'} = \begin{cases} c_{k'} & k' = 0, \ldots, K \\ 0 & k' = K + 1, \ldots, L - K - 1 \\ c_{k' - L + N} & k' = L - K, \ldots, L - 1. \end{cases} \tag{4}$$

A more densely sampled sequence $g(x_l)$, where $x_l = lX/L$ ($l = 0, \ldots, L - 1$), is now obtained by taking the inverse DFT of the sequence $d_{k'}$,

$$g(x_l) = \sum_{k'=0}^{L-1} d_{k'} \exp(j2\pi lk'/L), \tag{5}$$

for $l = 0, \ldots, L - 1$. ZP interpolation is generally viewed as a somewhat crude interpolation approach, but in fact nothing could be further from the truth. It can be shown that ZP interpolation is equivalent to CST interpolation, in that the

spatial-domain interpolation function corresponding to the ZP operation in frequency space is just $\sigma_N(x)$ given in Eq. 2 [9]. That is, it can be shown that

$$g(x_l) = \sum_{n=0}^{N-1} g(x_n)\sigma_N(x_l - x_n), \tag{6}$$

where $\sigma_N(x)$ is given by Eq. 2. This equivalence holds regardless of whether the Nyquist condition is satisfied. If it is satisfied, then ZP interpolation, like CST interpolation, is exact. Obviously, the CST interpolation formula can be used to estimate $g(x)$ at any arbitrary point x whereas zero-padding interpolation is constrained to interpolate onto a fixed, equispaced grid P times denser than the original samples. However, since ZP yields a continuous interpolated curve in the limit as $P \to \infty$, we shall treat these two approaches as one in the subsequent analysis, using the continuous form of Eq. 1 to represent them both.

The final periodic interpolation method to be examined is periodic spline (PS) interpolation. Splines are piecewise cubic polynomials that are continuous up to and including the second derivative at the joints between pieces [10, 11]. Periodic splines are further constrained to satisfy periodic boundary conditions. A spline $\widehat{g}(x)$ can be represented by

$$\widehat{g}(x) = a_n + b_n(x - x_n) + c_n(x - x_n)^2 + d_n(x - x_n)^3, \tag{7}$$

for $x \in [x_n, x_{n+1}]$, where the x_n are the abscissas at which the data is measured and $n = 0, \ldots, N - 1$. Fitting a spline is thus tantamount to finding the coefficients a_n, b_n, c_n, and d_n subject to interpolation, continuity, and boundary conditions. Because evaluating a spline at a particular point x involves evaluating Eq. 7 using the a_n, b_n, c_n, and d_n corresponding to the interval $[x_n, x_{n+1}]$ in which x falls, we can think of representing a spline as an N-component vector of functions, in which the nth component corresponds to the interval $[x_n, x_{n+1}]$ and contains a function of the form of Eq. 7 with the appropriate values substituted for a_n, b_n, c_n, and d_n. To reflect this understanding and simplify later calculations we introduce the somewhat unconventional notation

$$\widehat{\mathbf{g}}(x) = \mathbf{a} + \mathcal{D}(x)\mathbf{b} + \mathcal{D}^2(x)\mathbf{c} + \mathcal{D}^3(x)\mathbf{d}, \tag{8}$$

where $\widehat{\mathbf{g}}(x)$ is an N-element vector of functions, $\mathbf{a}$ is the N-element vector with coefficients a_n and likewise for $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$, and $\mathcal{D}$ is a diagonal matrix with $\mathcal{D}_{nn} = x - x_n$.

In fitting a spline, the coefficients are obtained by linear operations on the measured data. If the measured samples $g(x_n)$ are represented as an N-element vector $\mathbf{g}$, the vectors of coefficients can be found from $\mathbf{g}$ through matrix multiplications $\mathbf{a} = A\mathbf{g}$, $\mathbf{b} = B\mathbf{g}$, $\mathbf{c} = C\mathbf{g}$, and $\mathbf{d} = D\mathbf{g}$, where the matrices A, B, C, and D can be deduced from [11]. Substituting for $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$ in (8) in terms of these matrix products yields

$$\widehat{\mathbf{g}}(x) = \left[A + \mathcal{D}(x)B + \mathcal{D}^2(x)C + \mathcal{D}^3(x)D\right]\mathbf{g}. \tag{9}$$

While not theoretically exact for the interpolation of periodic, bandlimited functions, periodic splines can nonetheless be very accurate in that situation [12]. And in practice, when interpolating functions that are not exactly bandlimited, periodic splines often outperform CST and ZP interpolation, which are quite sensitive to departures from the bandlimited assumption, giving rise to high-frequency (Gibbs) artifacts [13]. In fact, if the smoothness of a function is defined in terms of the function's integrated second derivative, it can be shown that splines are the smoothest possible interpolant to any set of samples [14].

While questions about the accuracy of various interpolation approaches are paramount when the measured samples are known exactly, other concerns arise when the samples are known to be corrupted by noise. In particular, it becomes important to analyze how the noise is propagated into the interpolated samples. Consider now that the samples of $g(x)$ are corrupted by additive, zero-mean noise. These noisy samples can be represented as

$$g(x_n) = \langle g(x_n)\rangle + n(x_n), \tag{10}$$

where $g(x_n)$ and $n(x_n)$ are random variables, the latter representing zero-mean additive noise, and $\langle\rangle$ represents the expectation operator. The aim of this paper is to derive the covariance and variance of the curves interpolated by means of CST/ZP interpolation and PS interpolation, to compare these analytic predictions with results of Monte Carlo simulations, and to draw conclusions from these results about the suitability of the various approaches for the interpolation task encountered in few-view tomography.

II. METHODS

A. Analytic Derivations

Let $\widehat{g}(x)$ be a curve interpolated from the noisy samples of Eq. 10 by CST or ZP interpolation. The covariance between two points x and x' of this function is given by

$$\text{cov}(x, x') = \langle[\widehat{g}(x) - \langle\widehat{g}(x)\rangle]\,[\widehat{g}(x') - \langle\widehat{g}(x')\rangle]\rangle. \tag{11}$$

Of course, once the covariance has been computed, the variance at any point x is given by $\text{var}(x) = \text{cov}(x, x)$. Evaluation of Eq. 11 is straightforward as $\widehat{g}(x)$ is given by

$$\widehat{g}(x) = \sum_{n=0}^{N-1} \left(\langle g(x_n)\rangle + n(x_n)\right)\sigma_N(x - x_n), \tag{12}$$

where $\sigma_N(x)$ is given by Eq. 2, and thus

$$\text{cov}(x, x') = \left\langle \left[\sum_{n=0}^{N-1} n(x_n)\sigma_N(x - x_n)\right] \times \left[\sum_{m=0}^{N-1} n(x_m)\sigma_N(x' - x_m)\right]\right\rangle, \tag{13}$$

which can be rewritten

$$\text{cov}(x, x') = \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} \left[\sigma_N(x - x_n)\sigma_N(x' - x_m) \times \langle n(x_n)n(x_m)\rangle\right]. \tag{14}$$

Without additional assumptions about the noise, this expression cannot be simplified any further. In Section III we will consider simplifications to this expression for white noise and uncorrelated Poisson noise.

For periodic spline interpolation, the calculation of a covariance function is complicated by the fact that the points x and x' at which the covariance is being evaluated will in general fall into different intervals of the spline and the spline's value at these points will thus be specified by functions corresponding to different elements of the vector $\widehat{\mathbf{g}}(x)$. This is best handled by thinking in terms of computing an $N \times N$ matrix of covariance functions whose elements correspond to the possible combinations of pairs of intervals into which x and x' could fall. That is, given points x and x' between which it is desired to evaluate the covariance, one would compute the two intervals i and j in which these points fall and then evaluate the covariance function found in the ij^{th} position of the matrix. This matrix, which we denote with capital letters $\mathrm{COV}(x, x')$, is given by

$$\mathrm{COV}(x, x') = \left\langle [\widehat{\mathbf{g}}(x) - \langle \widehat{\mathbf{g}}(x) \rangle] [\widehat{\mathbf{g}}(x') - \langle \widehat{\mathbf{g}}(x') \rangle]^T \right\rangle, \quad (15)$$

or using Eq. 9

$$\mathrm{COV}(x, x') = \left\langle [A + \mathcal{D}(x)B + \mathcal{D}^2(x)C + \mathcal{D}^3(x)D] \right. \\ \left. \times \mathbf{n}\mathbf{n}^T [A + \mathcal{D}(x')B + \mathcal{D}^2(x')C + \mathcal{D}^3(x')D]^T \right\rangle, \quad (16)$$

where $\mathbf{n}$ is the N-element vector with entries $n(x_n)$, which of course reduces to

$$\mathrm{COV}(x, x') = [A + \mathcal{D}(x)B + \mathcal{D}^2(x)C + \mathcal{D}^3(x)D] \\ \times \left\langle \mathbf{n}\mathbf{n}^T \right\rangle [A + \mathcal{D}(x')B + \mathcal{D}^2(x')C + \mathcal{D}^3(x')D]^T. \quad (17)$$

Again, without further assumptions about the noise, this expression cannot be simplified any further. In Section III we will consider simplifications to this expression for white noise and uncorrelated Poisson noise.

B. Monte Carlo Simulations

To confirm the analytic results derived above and applied below to the cases of white and uncorrelated Poisson noise, we performed Monte Carlo simulations employing each of these kinds of noise. We first calculated analytically 128 equispaced samples of a typical periodic function encountered in tomography: the angular function corresponding to one bin of the sinogram of a Shepp-Logan head phantom [15]. This function is shown in Fig. 1. We generated $R=50{,}000$ noise realizations of these samples contaminated with additive, zero-mean Gaussian noise ($\sigma_o^2 = 16$) and 50,000 others with Poisson noise. We then interpolated each of these realizations to 1024 samples using the methods discussed above. For CST and ZP interpolation, we took $K = 63$, the largest value it could have while still being below the Nyquist frequency of the 128 samples. We calculated the sample variance $\widehat{\sigma}^2(x_j)$ at each interpolated point x_j using

$$\widehat{\sigma}^2(x_j) = \frac{1}{(R-1)} \sum_{i=1}^{R} \left\{ \widehat{g}_{(i)}(x_j) - \frac{1}{R} \sum_{i=1}^{R} \widehat{g}_{(i)}(x_j) \right\}^2, \quad (18)$$
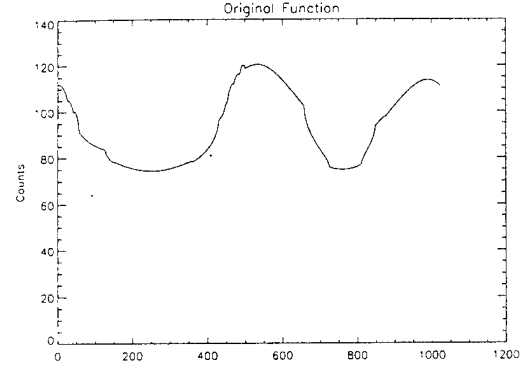

Figure 1: This periodic function, which is the angular function corresponding to one bin of a Shepp-Logan head phantom sinogram, was sampled 128 times and the samples then contaminated by Gaussian or Poisson noise prior to using CST/ZP and periodic spline interpolation to interpolate a sequence 8 times as dense.

where $\widehat{g}_{(i)}(x)$ is the curve interpolated from the ith noise realization. Similarly, we computed the sample covariance between each of 9 samples and the 1024 interpolated samples. This sample covariance between two points x_j and x_k was computed using

$$\widehat{C}(x_j, x_k) = \frac{1}{R-1} \sum_{i=1}^{R} \left\{ \widehat{g}_{(i)}(x_j) - \frac{1}{R} \sum_{i=1}^{R} \widehat{g}_{(i)}(x_j) \right\} \\ \times \left\{ \widehat{g}_{(i)}(x_k) - \frac{1}{R} \sum_{i=1}^{R} \widehat{g}_{(i)}(x_k) \right\}. \quad (19)$$

C. Image Noise Power Spectra

As mentioned in the introduction, periodic interpolation methods can be applied to the problem of interpolating additional projection views in few-view tomography. While the noise properties of the interpolated sinogram can be inferred from the analytic derivations discussed above, it is just as important to consider the noise properties of the images reconstructed from such sinograms, because these noise properties, as characterized by a noise power spectrum (NPS), greatly influence the detectability of small objects [16]. For this reason, we wished to calculate the NPS of images reconstructed after sinogram interpolation by CST/ZP and PS interpolation, as well as the NPS of images reconstructed directly from both a small and large number of views.

Though the noise power spectrum is strictly defined only for stationary noise processes, and the noise in images reconstructed from interpolated sinograms is not stationary due to the correlation between projections, there is precedent for examining the so-called average power spectrum of nonstationary processes [17]. To estimate the power spectrum for each approach, we generated 1000 sinograms containing only white, Gaussian noise ($\sigma_o^2 = 10$) and consisting of 128 projections bins and either 30 or 120 projection angles depending on which reconstruction approach was to be used. The sinograms were then either reconstructed directly by FBP or interpolated to 120 views using CST/ZP and PS

interpolation and then reconstructed by FBP. Each resulting noise image was multiplied by a circularly symmetric cosine window approximately 30 pixels in diameter and a 2D FFT computed. The resulting spectra were averaged and scaled so that the the volume under the power spectrum equaled the average variance in a circle of diameter 30 pixels.

III. RESULTS

A. White Noise

For CST and ZP interpolation, if the noise in the measured samples is assumed to be white, with constant variance σ_o^2 at every point, then in Eq. 14, $\langle n(x_n)n(x_m)\rangle = \sigma_o^2 \delta_{nm}$, where δ_{nm} is the Kronecker delta function and we may rewrite Eq. 14 as

$$\text{cov}(x, x') = \sigma_o^2 \sum_{n=0}^{N-1} \sigma_N(x - x_n)\sigma_N(x' - x_n). \quad (20)$$

This expression may be further simplified by comparing the sum with Eq. 1 and realizing that it can be viewed as a CST interpolation of the function $\sigma_N(x - x')$ sampled at points $x' = x_n$. Because $\sigma_N(x - x')$ is periodic and bandlimited to a frequency $K < N/2$, the interpolation is exact, and thus we conclude that

$$\text{cov}(x, x') = \sigma_o^2 \sigma_N(x - x'). \quad (21)$$

The covariance is thus seen to be shift invariant, depending only on the difference $x - x'$ between the positions of any two points in the interpolated curve and not on their absolute positions. From this result, the variance is easily obtained:

$$\text{var}(x) = \text{cov}(x, x) = \sigma_o^2 \sigma_N(0). \quad (22)$$

From the definition of $\sigma_N(x)$ in Eq. 2, we see that $\sigma_N(0) = (2K+1)/N$ and thus conclude that when applying CST and ZP interpolation to samples corrupted by white noise, the variance of the interpolated curve is constant everywhere and equal to $(2K+1)/N$ times the variance in the original samples. Because $N \geq 2K+1$, this factor is less than or equal to 1, with equality when N minimally satisfies the Nyquist condition.

The Monte Carlo studies support the analytic findings of Eqs. 21 and 22. Figure 2(a) depicts a portion of the 1024-point sample variance curve for CST and ZP interpolation. For comparison, the analytic prediction is shown in Fig. 2(b). The sample variance is seen to be approximately constant at all points, in agreement with the analytic prediction, and it is found to have average value 15.933, in agreement with the analytic prediction of 15.875. Figures 2(c) and 2(d) depict the sample and predicted covariance relative to a fixed point located midway between two measured samples. While they are not shown, the sample covariance curves relative to fixed points at other positions in the interval between measured points were found to have essentially identical shapes, as expected from the shift-invariant form of the analytic prediction.

For spline interpolation, the assumption of white noise with variance σ_o^2 implies that in Eq. 17, the outer product
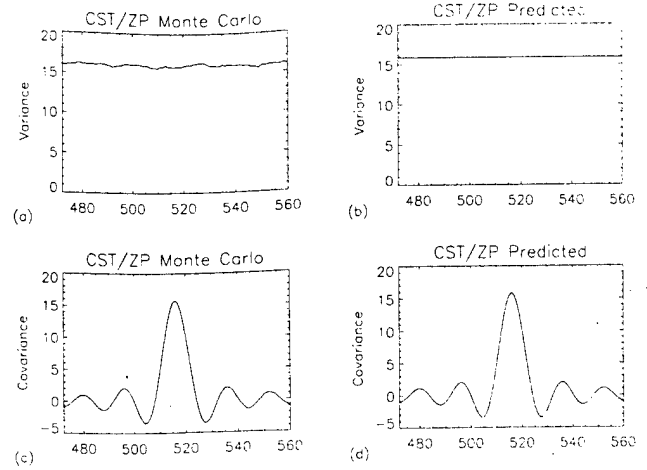

Figure 2: (a) Portion of sample variance curve for 50,000 realizations of CST/ZP interpolation from 128 samples to 1024 samples in the presence of additive, zero-mean Gaussian noise with $\sigma_o^2 = 16$. (b) Portion of analytically predicted variance curve for this task. (c) Portion of sample covariance curve relative to a fixed point midway between two measured points. (The shape of the curve for fixed points at other positions in the interval between measured samples was identical.) (d) Portion of analytically predicted covariance curve.
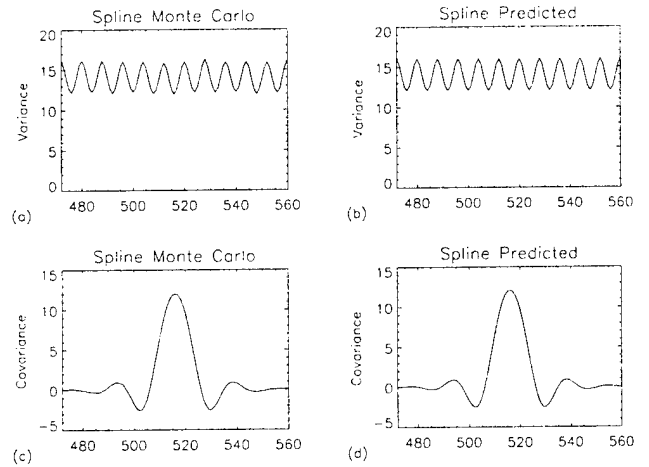

Figure 3: (a) Portion of sample variance curve for 50,000 realizations of periodic spline interpolation from 128 samples to 1024 samples in the presence of additive, zero-mean Gaussian noise with $\sigma_o^2 = 16$. (b) Portion of analytically predicted variance curve for this task. The maxima in these curves occur at the locations of the measured samples. (c) Portion of sample covariance curve relative to a fixed point midway between two measured points. (The shape of the curve for fixed points at other positions in the interval between measured samples was quite different.) (d) Portion of analytically predicted covariance curve for a fixed point midway between two measured points.

$\langle \mathbf{n}^T\mathbf{n}\rangle = \sigma_o^2 I$, where I is the identity matrix. The properties of the covariance expression given in Eq. 17 cannot be discerned by inspection and the result must be evaluated numerically. In particular, to obtain the variance of the interpolated function, the diagonal elements of the covariance matrix must be evaluated at $x = x'$. Figure 3(b) illustrates the result of such an evaluation. It is clearly seen that unlike CST and ZP

interpolation, the variance is not constant at interpolated points, but rather falls smoothly to a minimum at points midway between the original measured samples, while remaining unchanged at the measured points. The Monte Carlo results shown in Fig. 3(a) conform quite closely to this prediction. Figures 3(c) and 3(d) show a portion of the covariance function relative to a fixed point located midway between two measured samples. It is seen to have a slightly wider central lobe than the equivalent curve for CST/ZP interpolation but also to die out more quickly beyond this central lobe. Unlike in the CST/ZP case, the covariance curves relative to fixed points at other positions in the the interval between measured points were found to have quite different shapes, becoming asymmetric as the fixed point moves away from the center of the interval. Thus, the covariance of the interpolated curve is not shift-invariant for spline interpolation.

B. Poisson Noise

For CST and ZP interpolation, if the noise in the measured samples is assumed to be uncorrelated Poisson, then in Eq. 14, $\langle n(x_n)n(x_m)\rangle = \langle g(x_n)\rangle \delta_{nm}$, and we may rewrite Eq. 14 as

$$\text{cov}(x,x') = \sum_{n=0}^{N-1} \langle g(x_n)\rangle \sigma_N(x-x_n)\sigma_N(x'-x_n). \quad (23)$$

This sum can be viewed as a CST interpolation, this time of the function $\langle g(x')\rangle \sigma_N(x-x')$. However, this function, while periodic, is not in general bandlimited. However, if $\langle g(x')\rangle$ is slowly varying, at least over the interval between x and x', it can be assumed that the product is approximately bandlimited and thus that

$$\text{cov}(x,x') \cong \langle g(x')\rangle \sigma_N(x-x'). \quad (24)$$

The covariance function is seen to be approximately shift invariant, certainly as the fixed point is moved through any single interval between measured samples. The variance is again easily obtained:

$$\text{var}(x) = \text{cov}(x,x) \cong \langle g(x)\rangle \sigma_N(0). \quad (25)$$

Because $\sigma_N(0) = (2K+1)/N$, we find that the variance in the interpolated curve is flat locally (again, certainly between the positions of any two measured samples), while globally it tracks the variance in the measured samples.

The Monte Carlo studies confirm the analytic predictions of Eqs. 24 and 25. Figures 4(a) and 4(b) illustrate how the variance in the interpolated function, while locally constant, follows the variance in the measured samples over larger distances. The covariance function (not shown) is found to behave over short distances much as it did in the white noise case.

For spline interpolation of data contaminated by Poisson noise, the same general trends as in the white noise case are observed. The variance in the interpolated curve again falls smoothly to a minimum between measured samples, though now it tracks the variance in the measured samples globally.
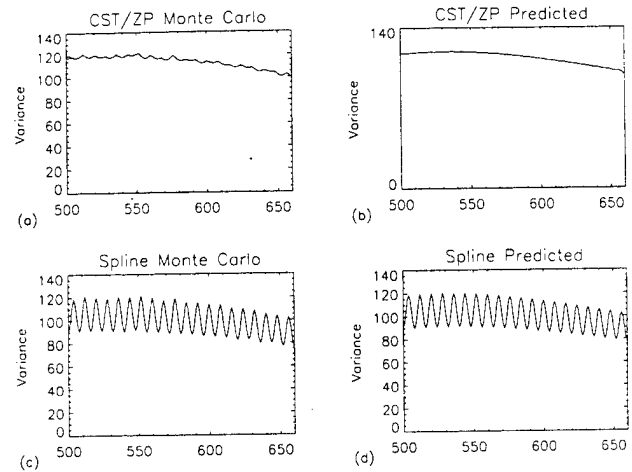

Figure 4: (a) Portion of a sample variance curve for 50,000 realizations of CST/ZP interpolation from 128 samples to 1024 samples in the presence of Poisson noise. (b) Portion of analytically predicted variance curve for this task. (c) Portion of a sample variance curve for 50,000 realizations of periodic spline interpolation from 128 samples to 1024 samples in the presence of Poisson noise. (d) Portion of analytically predicted variance curve for this task.
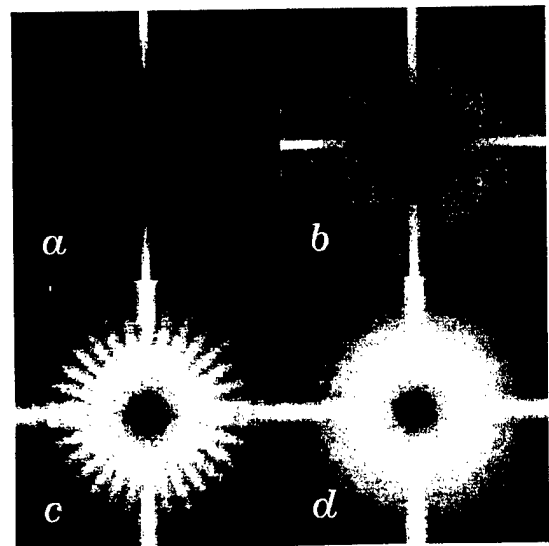

Figure 5: Power spectra for FBP reconstructions of stationary white Gaussian noise from (a) 30 projections. (b) 120 projections. (c) 120 projections interpolated by periodic spline interpolation from 30 projections. (d) 120 projections interpolated by zero-padding interpolation from 30 projections.

This is clearly seen in Figs. 4(c) and 4(d). The covariance curves were found to behave locally as they did in white noise case, changing form as the fixed point is swept through an interval between measured points.

C. Image noise power spectra

The spectra are shown in Fig. 5. The high-intensity streaks on the horizontal and vertical midlines are related to the use of linear interpolation in the radial direction during backprojection. The NPS for a direct reconstruction from 30 views shown in Fig. 5(a) is seen to consist of distinct rays

passing through the origin, as predicted by theory [18]. The NPS for a direct reconstruction from 120 views shown in Fig. 5(b) has a more continuous and circularly symmetric appearance. The NPS for reconstruction from 120 views interpolated from 30 views by PS interpolation (Fig. 5(c)) is seen to be quite non-uniform, a hybrid of the discrete spokes of the 30-angle reconstruction and the more uniform appearance of the 120-view reconstruction. The NPS for reconstruction from 120 views interpolated from 30 views by CST/ZP interpolation (Fig. 5(d)), on the other hand, is seen to resemble the 120-view NPS quite closely, being essentially continuous and circularly symmetric.

IV. DISCUSSION AND CONCLUSIONS

The differences between the CST/ZP and periodic spline interpolation approaches emerge most clearly in the white noise case discussed above. Here we saw that CST/ZP interpolation results in a curve with constant variance at all points (a factor of $(2K + 1)/N$ times the variance in the original measured samples), and having a covariance function that depends only on the distance $x - x'$ between two points. Put simply, if the noise in the measured samples is uncorrelated and wide-sense stationary, then the noise in the interpolated curve is also wide-sense stationary. Naturally, the noise does not remain uncorrelated. The situation is quite different for periodic spline interpolation. Here we saw that the variance remains equal to the variance in the original samples only at the interpolated points corresponding to those samples, and that it falls to a minimum at the midpoint between two such samples. Moreover, it was observed that the covariance function does not simply depend on the distance $x - x'$ between two points but depends also on the positions of the two points within their respective intervals between measured samples. Thus though the noise in the measured samples may be uncorrelated and wide-sense stationary, the noise in the curve interpolated by periodic spline interpolation is neither.

It is well known that when a stationary process is input to a linear, shift-invariant system, the output process is stationary as well [19]. Both CST/ZP and PS interpolation can be shown to involve linear, shift-invariant operations on the data, yet we have seen that given uncorrelated, stationary white noise in the measured samples, CST/ZP interpolation yields a stationary curve and PS interpolation does not. How do we explain this apparent paradox? The answer lies in the fact that the set of measured samples represents a discrete stationary process while the result of interpolation is a continuous process, and one must take care in comparing them. Systems theory makes no claim about the ability of a linear, shift-invariant system to turn a discrete stationary process into a continuous one. Indeed, if the original discrete process is viewed in the continuous domain it is not at all stationary: the variance falls to zero in the gaps between measured samples. What is truly surprising, then, is not that periodic spline interpolation fails to produce a stationary continuous process from the stationary discrete samples, but rather that CST/ZP interpolation succeeds in doing so.

The decision to use one approach or the other for the interpolation of additional angular views in few-view tomography should take into consideration the particular requirements of the study being performed. Given that the interpolatory accuracy of the two approaches is generally comparable, at first glance it would seem that the variance-reducing properties of the periodic spline interpolation might be preferable. However, this comes at the cost of having widely varying variance levels in neighboring projections, as well as non shift-invariant covariance. These non-uniformities do not noticeably affect the resulting image quality if one reconstructs directly from the interpolated sinogram and thus periodic spline interpolation may be appropriate if the images are only to be inspected visually, and particularly if the task involves the examination of relatively large structures. If, however, one seeks to perform any sort of principled smoothing on the projections or quantitation on the reconstructed image that requires knowledge of the statistical properties of the projections, then the stationarity-preserving properties of CST/ZP interpolation may be the best choice.

CST/ZP interpolation may also be the best choice if the task involves the detection of small objects. As mentioned in Section II, the noise properties of reconstructed images, as characterized by a noise power spectrum (NPS), greatly influence the detectability of small objects [16]. We examined the NPS of images reconstructed after sinogram interpolation by CST/ZP and PS and found the CST/ZP NPS to be much more similar to the NPS of images reconstructed from a full complement of 120 original views than was the spline NPS. Indeed, the spline NPS demonstrated the existence of non-uniform angular correlations that could potentially hinder detection tasks.

Finally, the ability of CST and ZP interpolation to reduce the variance in the interpolated curve by the factor $(2K + 1)/N$ relative to the variance in the measured samples deserves comment. This is simply an implicit exploitation of the ability to achieve noise reduction through oversampling and filtering [20, Ch. 5], and the mechanism is easiest to appreciate in the case of zero-padding interpolation for stationary white noise. The DFT of Eq. 3 has N terms, corresponding to frequencies out to $k = \pm N/2$ for N even or $k = \pm(N - 1)/2$ for N odd. In Eq. 4, however, we see that all DFT components corresponding to frequencies $|k| > K$ are explicitly set equal to zero, because $g(x)$ is assumed to be bandlimited (at least approximately) to frequency K. However, if the measured samples are corrupted by white noise, then their DFT can be seen as the sum of the true DFT and the DFT of the noise process, which has contributions at all N frequency components. Zeroing out all but $2K + 1$ of these would thus be expected to reduce the noise magnitude by the factor $(2K + 1)/N$.

V. ACKNOWLEDGMENT

VI. References

[1] R. A. Brooks, G. H. Weiss, and A. J. Talbert, "A new approach to interpolation in computed tomography," *J. Comput. Assist. Tomogr.*, vol. 2, pp. 577–585, 1978.

[2] H. Stark, J. W. Woods, I. Paul, and R. Hingorani, "Direct Fourier reconstruction in computer tomography," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 29, pp. 237–245, 1981.

[3] H. Stark, "Sampling theorems in polar coordinates," *J. Opt. Soc. Am.*, vol. 69, pp. 1519–1525, 1979.

[4] H. Stark and M. Wengrovitz, "Comments and corrections on the use of polar sampling theorems in CT," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 1329–1331, 1983.

[5] R. W. Schafer and L. R. Rabiner, "A digital signal processing approach to interpolation," *Proc. IEEE*, vol. 61, pp. 692–702, 1973.

[6] K. P. Prasad and P. Satyanarayana, "Fast interpolation algorithm using FFT," *Electron. Lett.*, vol. 22, pp. 185–187, 1986.

[7] D. Fraser, "Interpolation by the FFT revisited—an experimental investigation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 665–675, 1989.

[8] T. J. Cavicchi, "DFT time-domain interpolation," *IEE Proceedings-F*, vol. 139, pp. 207–211, 1992.

[9] P. J. La Rivière and X. Pan, "Mathematical equivalence of zero-padding and circular sampling thorem interpolation with implications for direct Fourier image reconstruction," in *Proc. SPIE*, vol. 3338, pp. 1117–1126, 1998.

[10] H. Spath, *One-Dimensional Spline Interpolation Algorithms*. Wellesley, MA: A. K. Peters, 1993.

[11] C. H. Reinsch, "Smoothing by spline functions," *Numer. Math.*, vol. 10, pp. 177–183, 1967.

[12] P. J. La Rivière and X. Pan, "Comparison of angular interpolation approaches in few-view tomography using statistical hypothesis testing," in *Proc. SPIE*, vol. 3661, 1999. (To appear).

[13] E. O. Brigham, *The Fast Fourier Transform and Its Applications*. Englewood Cliffs: Prentice Hall, 1988.

[14] P. J. Green and B. W. Silverman, *Nonparametric Regression and Generalized Linear Models*. London: Chapman Hall, 1994.

[15] L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," *IEEE Trans. Nucl. Sci.*, vol. 21, pp. 21–43, 1974.

[16] International Comission on Radiation Units and Measurements, "Medical imaging: The assessment of image quality," 7910 Woodmont Ave., Bethesda, MD, 20814, 1996.

[17] S. C. Moore, M. F. Kijewski, S. P. Müller, and B. L. Holman, "SPECT image noise power: Effects of nonstationary projection noise and attenuation compensation," *J. Nucl. Med.*, vol. 29, pp. 1704–1709, 1988.

[18] K. Hanson, "Detectability in computed tomographic images," *Med. Phys.*, vol. 6, pp. 441–451, 1979.

[19] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw Hill, 1991.

[20] R. J. Marks, *Introduction to Shannon Sampling and Interpolation Theory*. New York: Springer Verlag, 1991.

Few-View Tomography Using Roughness-Penalized Nonparametric Regression and Periodic Spline Interpolation

P. J. La Rivière, *Student Member, IEEE*, and X. Pan, *Member, IEEE*

Department of Radiology, The University of Chicago, Chicago, IL, 60637

Abstract

The ability to reconstruct high-quality tomographic images from a smaller number of projections than is usually used could reduce imaging time for many nuclear-medicine studies. This would particularly benefit studies such as cardiac SPECT where patient motion during long acquisitions can lead to motion artifacts in the reconstructed images. To this end, we have investigated sinogram pre-processing techniques designed to enable filtered backprojection (FBP) to produce high-quality reconstructions from a small number of views. Each projection is first smoothed by performing roughness-penalized nonparametric regression using a generalized linear model that explicitly accounts for the Poisson statistics of the data. The resulting fit curves are natural cubic splines. After smoothing, additional angular views are generated using periodic spline interpolation, and images are reconstructed using FBP. The algorithm was tested on data from SPECT studies of a cardiac phantom placed at various radial offsets to enable examination of the algorithm's dependence on the radial extent of the object being imaged.

I. INTRODUCTION

In routine nuclear-medicine tomographic studies, there is usually a tradeoff between image quality and imaging time. Increasing the number of angular views acquired, the number of counts per view, or both will generally improve image quality but will also lengthen imaging time. In general, concerns about image quality take precedence over concerns about imaging time. However, in studies where the patient must assume an awkward or uncomfortable position, long acquisition times can potentially lead to patient motion and thus to motion artifacts in the reconstructed images. One way to reduce imaging time without significant sacrifice of image quality is to use a continuous acquisition mode, in which the time wasted in moving the camera between views is eliminated. However, continuous acquisition is not recommended for some of the studies most plagued by motion artifacts, such as gated cardiac SPECT. In such studies, reducing imaging time must be accomplished either by reducing the number of angular views or by reducing the number of counts per view. In this paper, we focus on the first approach, investigating algorithms tailored to generate diagnostically useful images from a smaller number of angular views than is usually used while holding the number of counts per view constant.

One strategy for reconstructing acceptable images in few-view tomography is to incorporate into the reconstruction process as much prior information as possible about the expected image. For instance, constraints regarding the size [1], symmetry properties [2], or even the mean [3] of the expected image may be used. However, this approach generally leads to iterative algorithms involving multiple reconstruction and reprojection steps that are computationally intensive and subject to concerns regarding convergence and regularization. Moreover, the incorporation of prior information may involve a subjective and time-consuming operator-dependent step. These drawbacks are generally well justified by the resulting image quality in extreme situations involving reconstruction of a fairly complex object from 10 or fewer projections, where it is hardly possible to generate acceptable images without incorporating prior information. However, our interest is in more moderate situations, such as reconstruction from 60 or 30 views when 120 might normally be used, and in these situations, a more computationally efficient, fully automatic, robust algorithm may be preferable. For these reasons, we focus in this work on the development of algorithms for few-view reconstruction involving sinogram pre-processing followed by reconstruction by filtered backprojection (FBP).

In the absence of additional *a priori* constraints about the object being imaged, the minimum number of angular views required to produce an accurate tomographic reconstruction of a given object using FBP is dictated by two factors. First, the angular sampling of the object's sinogram must satisfy the Nyquist sampling condition. Absent this, any reconstruction is doomed to suffer from angular aliasing artifacts. Second, the number of angular samples must satisfy FBP's implicit assumptions about the density of angular sampling. It is well known that FBP reconstructions from a small number of angular views are degraded by prominent star-shaped artifacts. These two conditions are not in general equivalent, as can be appreciated most keenly when considering the case of imaging a circularly symmetric object. In this instance, a single projection view is sufficient to satisfy the Nyquist sampling condition, while a FBP reconstruction from this single view would be an uninterpretable set of parallel streaks. Indeed, Brooks *et al.* have shown that for a circularly symmetric object imaged with N bins per projection, a minimum of $\sim 1.1\pi N/4$ projections must be acquired over 180° to produce an essentially artifact-free reconstruction using FBP [4].

The Nyquist condition is the more fundamental of the two sampling requirements discussed above, because when it is satisfied by a number of samples less than the number required by the reconstruction algorithm, it is in principle possible to interpolate exactly the additional views needed. This is only strictly true in the absence of noise, of course, a condition that rarely obtains in emission tomography. In the presence of noise, angular interpolation usually leads to severe circular artifacts, because the noise process is rarely if ever bandlimited to the Nyquist frequency of the samples.

However, by smoothing each projection prior to interpolation, the noise-engendered interpolation artifacts can be effectively eliminated. While simple linear, shift-invariant smoothing filters, such as a Hanning window, lead to reasonable results, we opted to explore a more sophisticated smoothing approach that makes better use of the information contained in the data. The approach we chose was roughness-penalized nonparametric regression using an explicit Poisson statistical model [5], which leads to fit functions that are natural cubic splines, piecewise cubic polynomials that are continuous up to and including the second derivative and satisfying the so-called natural boundary conditions. While this approach is similar in spirit to the information-weighted splines of Fessler [6], the use of the explicit Poisson model in the present case leads to a rather different, iterative algorithm. Moreover, rather than choosing the smoothing parameter *a priori*, we have implemented an automatic algorithm for determining it, based on the principle of cross-validation and adapted appropriately for Poisson-distributed data. The subsequent interpolation of additional views is also performed using splines, this time satisfying periodic boundary conditions. Image reconstruction then proceeds as usual, with due recognition of the fact that the sinogram has already been smoothed.

II. METHODS

A. Interpolation of angular views

Consider a tomographic acquisition that yields a two-dimensional discrete sinogram $p(\xi_n, \phi_m)$, where ξ_n, $n = 1, \ldots, N$, is the projection bin and ϕ_m, $m = 1, \ldots, M$, is the projection angle. The number M of angular samples is assumed to satisfy the Nyquist condition at least approximately (i.e., the energy of the angular spectrum beyond the Nyquist frequency is assumed to fall below some small threshold). We wish to increase the number of angular samples to KM, where K is an integer and KM is large enough to eliminate star artifacts from an FBP reconstruction, by interpolating additional views between the measured ones. To do this, the sinogram is viewed as a set of 1D sampled functions of projection angle, each labeled by a projection bin ξ_n, with the samples denoted by $p_{\xi_n}(\phi_m)$. Then continuous one-dimensional (1D) interpolating curves $p_{\xi_n}^i(\phi)$, where the superscript i indicates interpolated, are fit to each of these sampled functions and resampled to obtain the additional angular views.

Ideally, a periodic interpolation method should be chosen in order to make use of the inherent periodicity of the angular samples. In image reconstruction, periodic interpolation has primarily been studied in the context of direct Fourier reconstruction, where there is a need to interpolate from a polar to a Cartesian grid in the Fourier space of the image function while exploiting the periodicity of the azimuthal samples [7–9]. Fewer studies exist of interpolation in sinogram space [1, 10], though many of the Fourier-space interpolation techniques can be readily applied in sinogram space, including simple schemes such as nearest-neighbor

and linear interpolation with periodic boundary conditions as well as the more complex circular sampling theorem (CST) and zero-padding approaches [11–14]. Despite its simplicity, zero-padding interpolation is quite accurate; indeed, it can be shown [15] that it is mathematically equivalent to CST interpolation and that both are exact when interpolating bandlimited, periodic functions whose samples satisfy the Nyquist condition. Because ZP interpolation exploits the efficient FFT algorithm, it has a much lower computational burden than CST interpolation, and is thus to be preferred in this context.

Despite the theoretical exactness of CST/ZP interpolation when the Nyquist condition is satisfied, the two methods are quite sensitive to mild violations of the criterion, giving rise to the well-known Gibbs oscillations that radiate widely from sharp edges [16]. This shortcoming prompted us to explore yet another interpolation technique based on cubic splines that while not theoretically exact is known to be very accurate and quite robust in the face of violations of the Nyquist condition. Our confidence in the use of splines was bolstered by studies we performed [17] of 1D interpolation in which samples of known analytic functions representing the angular variations of projections of Shepp-Logan-like and breast phantoms [18] were interpolated at intermediate points and the interpolated values then compared to the known values at those points. In these tests, spline interpolation was found to have a statistically significant lower root-mean-squared error than linear interpolation or CST/ZP interpolation for virtually all sampling intervals.

A periodic cubic spline is a curve comprised of generally different third-order polynomials between each pair of known abscissas ϕ_m and ϕ_{m+1}, with the overall curve being continuous up to and including the second derivative at each abscissa [19]. Naturally, in the case of an interpolating spline (as opposed to a smoothing spline), the curve $p_{\xi_n}^i(\phi)$ is also constrained to pass through the known ordinate values $p_{\xi_n}(\phi_m)$ at each abscissa ϕ_m. The spline can be represented as

$$p_{\xi_n}^i(\phi) = a_m + b_m\phi + c_m\phi^2/2 + d_m\phi^3/3. \tag{1}$$

for $\phi \in [\phi_{m}, \phi_{m+1}]$, where $m = 1, \ldots, M$. The process of fitting an interpolating spline is then tantamount to solving a set of linear equations for the defining coefficients a_m, b_m, c_m, and d_m in each interval $[\phi_m, \phi_{m+1}]$ such that the continuity and interpolation conditions are satisfied. For a periodic spline, the coefficients must also satisfy periodic boundary conditions.

In order to illustrate the effects of this interpolation and to motivate our approach to mitigating noise, we have reconstructed images from simulated projections of a numerical phantom. The true phantom is shown in Fig. 1(a). In Fig. 1(b), the phantom is shown reconstructed by ramp-filtered FBP from 120 noiseless angular views. In Fig. 1(c), a reconstruction from 30 views, we observe the star-shaped artifacts discussed in the introduction. Figure 1(d) illustrates the results of ramp-filtered FBP reconstruction of the phantom after interpolating from 30 to 120 angular views using periodic spline interpolation. We see that the star-shaped artifacts have
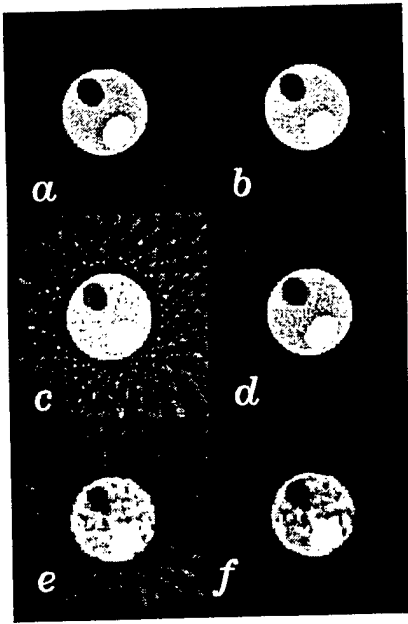
Figure 1: Reconstructions of a numerical phantom. (a) True phantom. (b) Reconstruction by ramp-filtered FBP from 120 noiseless angular views. (c) Reconstruction by ramp-filtered FBP from 30 noiseless angular views. (d) Reconstruction by ramp-filtered FBP after interpolation from 30 to 120 noiseless angular views. (e) Reconstruction by Hanning-filtered FBP from 30 angular views with Poisson noise. (f) Reconstruction by Hanning-filtered FBP after interpolation from 30 angular views with Poisson noise to 120 views.

been essentially eliminated. In Fig. 1(e), we show the results of a Hanning-filtered reconstruction from 30 angular views with Poisson noise (125,000 total counts). The star-shaped background artifacts are still present and we observe the image to be somewhat noisy. Finally, in Fig. 1(f), we show the results of a Hanning-filtered reconstruction of the phantom after interpolating from 30 to 120 views. We observe that the star-shaped artifacts have been eliminated as before, but that circular noise artifacts have appeared in the image.

B. Smoothing

While the results of spline interpolation in the noiseless case (Fig. 1(d)) are encouraging, the poor reconstruction in the noisy case (Fig. 1(f)) indicates that prior smoothing of the sinogram may be necessary if spline interpolation of angular views is to succeed in the presence of noise. Rather than simply apply a linear, shift-invariant filter to the noisy projection data, we decided to explore a more principled statistical approach using roughness-penalized nonparametric regression based on a generalized linear model (GLM) [5] that explicitly accounts for the Poisson nature of the SPECT data.

1) Nonparametric regression using a GLM

Regression analysis with a single explanatory variable considers the problem of fitting a curve to a set of data pairs (Y_i, x_i), $(i = 1, \ldots, N)$, where the Y_i are the measured values of the quantity of interest and the x_i the corresponding values of the explanatory variable. The variation in the Y_i is assumed to have two components: a systematic component captured by a vector of predictors θ_i that depends on the x_i, and a

random component specifying the distribution of the Y_i given θ_i. In classical linear regression, for example, the systematic component is assumed to be of the form $\theta_i = ax_i + b$, and the Y_i are assumed to be normally distributed about the θ_i, i.e., $Y_i \sim N(\theta_i, \sigma^2)$. In general, maximum-likelihood methods are used to estimate regression curves, and in the case of classical linear regression, the maximum-likelihood estimates of a and b are easily shown to be those that minimize the sum of squares $\sum_{i=1}^{N} (Y_i - ax_i - b)^2$ [20].

Nonparametric regression using a GLM relaxes both of the assumptions of classical linear regression. First, it eliminates the assumption that the predictors θ_i depend on the explanatory variable in a simple parametric way, representing them instead as an arbitrary function of the explanatory variables: $\theta_i = g(x_i)$. This would seem to complicate the problem immensely, turning a relatively straightforward finite-dimensional estimation problem into an intractable infinite-dimensional one. However, in practice the problem is made tractable by adding the further constraint that the estimated curve $g(x)$ be smooth and enforcing this constraint by penalizing the likelihood with a term of the form $\int g''(x)^2 dx$. If the unpenalized likelihood depends on $g(x)$ only through its values $g(x_i)$ at the measured points x_i, $i = 1, \ldots, N$ (which is usually the case), it can be shown that the minimizer of the penalized likelihood is necessarily a natural cubic spline [5]. These are simply cubic splines that are constrained to be linear outside the set of measured points, and like the periodic splines discussed above, they can be specified by a finite number of coefficients. (In this sense, roughness-penalized nonparametric regression does yield a parametrized curve, but given that there are at least as many parameters as observations, the spirit is rather different from that of standard parametric techniques.)

Nonparametric regression using a GLM also relaxes the assumption that the data are normally distributed in favor of the much broader class of exponential distributions, which have probability densities of the form

$$p(y_i \mid \theta_i, \phi) = \exp\left(\frac{y_i \theta_i - b(\theta_i)}{\phi} + c(y_i, \phi)\right), \quad (2)$$

where θ_i is the so-called natural parameter of the exponential family and ϕ is a scale parameter [21]. This family comprises many well-known distributions, each corresponding to a different choice of the functions b and c. Of particular interest to emission tomography is the choice $b(\theta_i) = e^{\theta_i}$, $\phi = 1$, $c(y_i, \phi) = -\log(y_i!)$, which corresponds to a Poisson distribution with parameter $\lambda_i = e^{\theta_i}$. If a nonparametric dependence $\theta_i = g(x_i)$ of the predictor on the explanatory variables is assumed, the log-likelihood of N independent observations Y_i drawn from a Poisson density is given by

$$\ell(g, \phi) = \sum_{i=1}^{N} (Y_i g(x_i) - \exp[g(x_i)] - \log(Y_i!)). \quad (3)$$

The goal of roughness-penalized nonparametric regression is to estimate the curve $g(x)$ that maximizes this log-likelihood

subject to the penalty $\int g''(x)^2 dx$, i.e. to maximize

$$\sum_{i=1}^{N} \{Y_i g(x_i) - \exp[g(x_i)]\} - \frac{1}{2}\alpha \int g''(x)^2 dx, \quad (4)$$

where terms independent of g have been dropped and where α is the so-called smoothing parameter, to be discussed below. Note that $g(x)$ here represents an estimate of the *log* of the Poisson parameter $\lambda(x)$. As mentioned above, this expression can be shown to be maximized by a natural cubic spline and it can also be shown that for natural cubic splines, the penalty $-\frac{1}{2}\int g''(x)^2 dx = -\frac{1}{2}\alpha g^T K g$, where g is an N-element vector with $g_i = g(x_i)$, and K is an $N \times N$ matrix determined by the spacing of the measurement points x_i. Because the natural cubic spline interpolating any specified set of points $g(x_i)$ is unique, finding $g(x)$ is thus tantamount to finding g [5].

As is traditional in GLMs [21], we use Fisher scoring to find the g that maximizes Eq. 4, which yields the following iterative update equation

$$g^{(k+1)} = (W + \alpha K)^{-1} W z^{(k)}, \quad (5)$$

where z is an N-element vector with, for Poisson data, components

$$z_i^{(k)} = g_i^{(k)} + \frac{Y_i - \exp\left(g_i^{(k)}\right)}{\exp\left(g_i^{(k)}\right)}, \quad (6)$$

W is a diagonal matrix with, for Poisson data, entries $W_{ii} = \exp\left(g_i^{(k)}\right)$, and the superscript (k) refers to the kth iteration. The initial estimate $g^{(0)}$ is chosen to have components $g_i^{(0)} = \log\{\max(Y_i, \epsilon)\}$, where ϵ is a small positive constant introduced to avoid computing $\log(0)$. Iteration continues until the sum of the absolute changes in the components of g from one iteration to the next falls below a prespecified threshold. While this would seem to be a very computationally intensive procedure, the banded structure of some of the matrices involved can be exploited to keep the algorithm to $O(N)$.

2) Choice of the smoothing parameter

The choice of the smoothing parameter α profoundly influences the appearance of the fit curve $g(x)$, for α determines the relative influence of the two terms in the penalized-likelihood expression, the first rewarding goodness-of-fit to the data, the second rewarding smoothness. A small value of α leads to a ragged curve while a large value of α leads to a smooth curve. In most applications, a value between these two extremes is desirable, and while this can be found through trial and error for most datasets, a more principled and automatic approach would clearly be preferred.

One such automatic approach to choosing the smoothing parameter is based on the principle of cross validation (CV), which has been discussed in the context of image processing by Galatsanos and Katsaggelos [22]. The approach is grounded in the assumption that the choice of α should yield a fit curve $g(x)$ that accurately predicts the outcomes of further observations. Remarkably, the predictive accuracy of the fit curve can be quantified solely on the basis of the fit values and the measured data [23]. This so-called CV score can be expressed as

$$CV(\alpha) = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{Y_i - g(x_i)}{1 - A_{ii}(\alpha)}\right\}^2, \quad (7)$$

where A_{ii} are the diagonals of the hat matrix A, which links the values of the estimate at the measured points to the values of the observations at those points: $g = A(\alpha)Y$. The value of α minimizing this curve is considered optimal, and can generally be found fairly quickly using a golden section search minimization approach.

The CV score of Eq. 7, based on a residual sum of squares, is more appropriate for normally distributed data than for the more general class of distributions encompassed by GLMs. For GLMs, we follow O'Sullivan *et al.* [24], who propose replacing the residual sum of squares with the generalized Pearson χ^2 statistic, which for Poisson distributed data is given by

$$CV_{\text{GLM}}(\alpha) = \frac{1}{N}\sum_{i=1}^{N}\frac{(Y_i - \exp(g_i))^2/\exp(g_i)}{(1 - A_{ii}(\alpha))^2}. \quad (8)$$

where the A_{ii} are the diagonal elements of a matrix A given by $(W + \alpha K)^{-1}W$, where W is as above, evaluated after the final iteration. Calculating the CV score would seem to be an $O(N^2)$ operation, for computing the matrix A involves inverting the matrix $(W + \alpha K)$. However, two facts work in our favor: first, $(W + \alpha K)$ can be manipulated to take advantage of band structures and second, only the diagonal elements of A are needed. Hutchinson and de Hoog [25] have developed an algorithm that allows the diagonal elements of the inverse of a band matrix to be computed in $O(N)$, and this is the approach we have used.

C. Data Acquisition and Processing

In order to examine the response of the algorithm to increasingly difficult interpolation tasks, we imaged a compact object placed at increasing radial offsets. This increases the bandwidth of the angular functions at each projection bin. Specifically, we acquired projections of a Data Spectrum ventricular phantom placed at five different radial offsets from the COR: 0, 5, 9, 12, and 15 cm. The phantom was filled with 121 MBq (3.27 mCi) of Tc-99m, contained a 1-cm defect insert, and was not placed within a water-filled torso phantom. We imaged this phantom with a Picker 3000XP three-headed SPECT system fit with low-energy, high-resolution, parallel-hole collimators, acquiring studies containing 120 angular views over 360°. We used a 25-cm radius circular orbit and step-and-shoot mode for all of the acquisitions; each head acquired to a 128x128 pixel image, though we preserved only the 32 slices spanning the phantom. A total of about 500,000 counts was collected. From this data we extracted 3D sinograms corresponding to 15, 30, 60, and 120 views, respectively. Thus we had 20 different sinograms,

corresponding to the 20 possible combinations of radial offset and number of angular views. We reconstructed images from these 20 sinograms using four different processing techniques:

1. No pre-smoothing of the sinogram and slice-by-slice reconstruction from available views by FBP using a Hanning filter (cutoff=0.8).

2. No pre-smoothing of the sinogram, spline interpolation from the available views to 120 angular views, and slice-by-slice reconstruction by FBP using a Hanning filter (cutoff=0.8).

3. Roughness-penalized nonparametric regression smoothing of the sinogram and slice-by-slice reconstruction from the available views by FBP using a ramp filter (cutoff=1.0).

4. Roughness-penalized nonparametric regression smoothing of the sinogram, spline interpolation from the available views to 120 views, and slice-by-slice reconstruction by FBP using a ramp filter (cutoff=1.0).

The particular application of the nonparametric regression technique to this data requires some explanation. The projection data for each of the radial offsets is a 3D sinogram of 128 bins, 32 slices, and 120 angles. While the smoothing technique was applied to each projection in each slice independently, we decided it would be wise to use the same smoothing parameter for all the projections. To do otherwise would have invited inconsistencies in the smoothed sinogram likely to produce streaks and other artifacts in the reconstructed images. At the same, we wished to select the smoothing parameter to be applied to the data using some form of cross-validation. The solution was to string together all of the 1D projections in the complete 3D sinogram into a single, long 1D function, and to find the smoothing parameter that minimized the CV score for that function when smoothed by the roughness-penalized nonparametric technique. This value of α was then used in smoothing each of the projections in this sinogram individually.

III. RESULTS

For ease of comparison and simplicity of presentation, we have grouped the reconstructed images by the number of angles in the original sinogram, and we show in Fig. 2 the results for only three of the radial offsets: 0, 9, and 15 cm. These three suffice to illustrate the trends observed. For each combination of number of angles and radial offset we show the results of reconstructing by use of each of the four techniques outlined above.

We observe that reconstructions from available views without pre-smoothing or interpolation display star-shaped artifacts and a mottled appearance when the number of views is small. Interpolation alone mitigates the star-shaped artifacts but leads to severe circular artifacts, particularly in the case of a small number of views and a large radial offset. Smoothing alone reduces the noise visibility but has little effect on the star-shaped artifacts. The combination of smoothing and interpolation still produces circular interpolation artifacts in the case of a large radial offset combined with a small number of original views, but these are less severe than when interpolation alone is used. Overall, though, visually appealing reconstructions result for less challenging combinations of radial offset and number of views, including as few as 15 angles in the 0-cm offset case.

While Fig. 2 demonstrates that the algorithms described can produce visually satisfactory reconstructions of the cardiac phantom from relatively small numbers of views, the most critical question is whether this can be achieved without hindering the detection of the small perfusion defects that is usually the goal of cardiac SPECT imaging. To answer this question we generated bullseye plots [26] from each set of reconstructions. To construct these plots, one starts with a set of short-axis slices of the left ventricle, which have the appearance of annuli. Each slice is mapped to a different ring of a dartboard-like grid divided into radial and azimuthal sectors, with the apex corresponding to the innermost ring (the bullseye) and the base corresponding to the outermost ring. The value of the integrated activity in a finite angular sector of each annulus is mapped to the appropriate azimuthal sector of the dartboard. If the reconstruction of the ventricle contained a uniform distribution of activity, then the bullseye plot would be uniform. But if the activity were non-uniform, for instance if there were a perfusion defect having lower activity than surrounding areas, then the appropriate sector of the dartboard would appear darker than the surrounding area. The bullseye plots corresponding to reconstructions from few-view sinograms that have been processed by the spline smoothing and interpolation techniques are shown in Fig. 3, along with bullseye plots corresponding to 120-angle sinograms reconstructed without spline processing. Our phantom contained a 1-cm perfusion defect insert, which produces a depression that is well resolved in many of the bullseye plots.

It is clear from these bullseye plots that the defect remains detectable for as few as 15 angles in the case of the 0-cm offset, as few as 30 angles in the case of the 9-cm offset, and as few as 60 angles in the case of the 15-cm offset. These findings correlate well with the visual appearances of the images in Fig. 2. The plots in which the defect is not visible correspond to the images in which severe interpolation artifacts are evident.

IV. DISCUSSION AND CONCLUSIONS

We have presented a sinogram pre-processing technique combining spline-based smoothing and interpolation that enables FBP to produce high-quality tomographic reconstructions from a smaller number of views than is usually used. The technique is applicable to situations where the number of views needed to satisfy the Nyquist condition on the sampling of the angular part of the sinogram is less than the number of views required by FBP to produce artifact-free images. In this situation, we first smooth each 1D projection
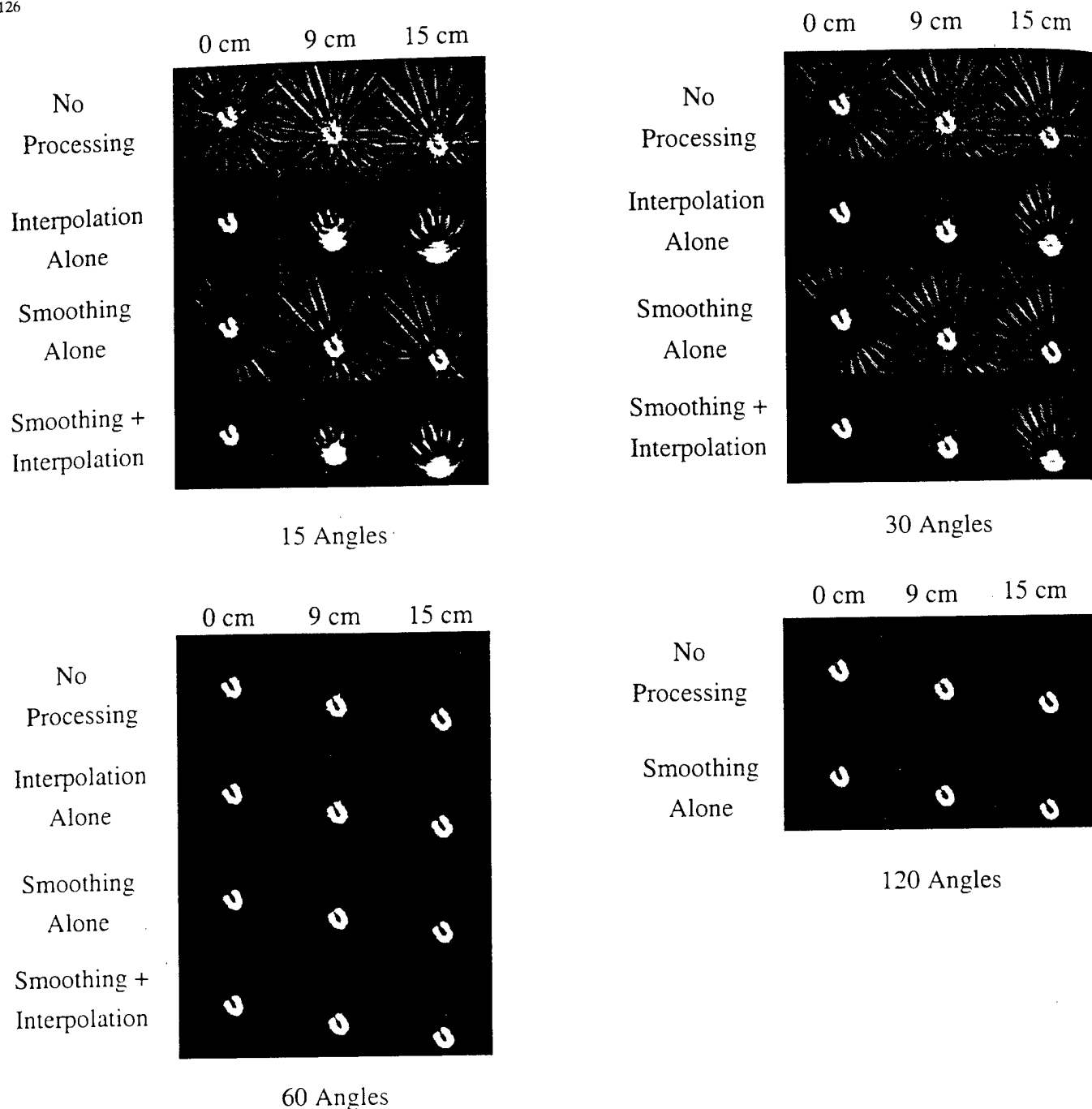
0 cm   9 cm   15 cm

No Processing

Interpolation Alone

Smoothing Alone

Smoothing + Interpolation

15 Angles

0 cm   9 cm   15 cm

No Processing

Interpolation Alone

Smoothing Alone

Smoothing + Interpolation

30 Angles

0 cm   9 cm   15 cm

No Processing

Interpolation Alone

Smoothing Alone

Smoothing + Interpolation

60 Angles

0 cm   9 cm   15 cm

No Processing

Smoothing Alone

120 Angles

Figure 2: A representative slice of a cardiac phantom imaged at three different radial offsets and reconstructed with use of four different pre-processing approaches from 15, 30, 60, and 120 projection angles.

using a roughness-penalized nonparametric regression approach based on a GLM that explicitly models the Poisson statistics of the measured data. One-dimensional periodic interpolation splines are then fit in the angular direction and resampled to generate additional views. Ramp-filtered FBP is then applied to the interpolated sinogram.

Because it is not possible to derive closed-form expressions for the Nyquist and FBP sampling requirements of non-circularly symmetric objects, we have tested the limits of the algorithm by conducting an experimental investigation performing reconstructions from various numbers of projections of a cardiac phantom placed at various radial offsets from the center of rotation of a SPECT system. The ability of the algorithm to produce visually appealing images that still capture small perfusion defects breaks down for large radial offsets combined with small numbers of starting views. This corresponds to the situation when the number of initial angular samples strongly fails to satisfy the Nyquist condition for some or all projection bins and thus the assumptions underlying the approach are violated.

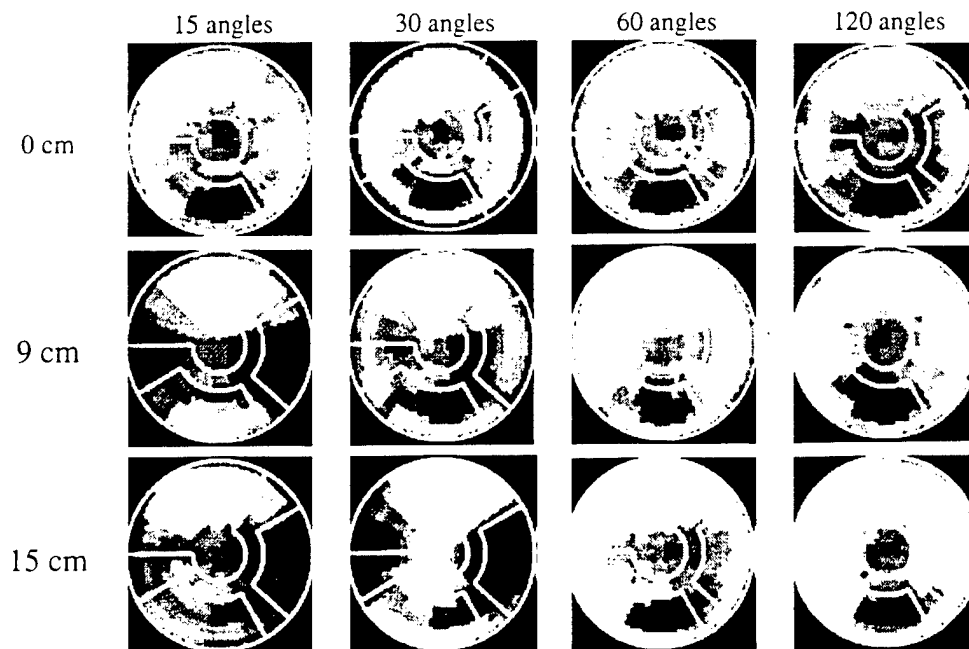| | 15 angles | 30 angles | 60 angles | 120 angles |

Figure 3: Bulleye plots for the cardiac phantom placed at various radial offsets and for four different numbers of projection angles. The reconstructions from 15, 30, and 60 angles used the nonparametric regression smoothing technique followed by periodic spline interpolation to 120 views prior to reconstruction by ramp-filtered FBP. The reconstructions from 120 angles simply entailed Hanning-filtered FBP.

V. ACKNOWLEDGMENT

VI. REFERENCES

[1] P. B. Heffernan and R. H. T. Bates, "Image reconstruction from projections. VI: Comparison of interpolation methods," *Optik*, vol. 60, pp. 129–142, 1982.

[2] R. A. Crowther, D. J. DeRosier, and A. Klug, "The reconstruction of a three-dimensional structure from projections and its application to electron microscopy," *Proc. Roy. Soc. Lond A*, vol. 317, pp. 319–340, 1970.

[3] K. A. Hanson and G. W. Wecksung, "Bayesian approach to limited-angle reconstruction in computed tomography," *J. Opt. Soc. Am.*, vol. 73, pp. 1501–1509, 1983.

[4] R. A. Brooks, G. H. Weiss, and A. J. Talbert, "A new approach to interpolation in computed tomography," *J. Comput. Assist. Tomogr.*, vol. 2, pp. 577–585, 1978.

[5] P. J. Green and B. W. Silverman, *Nonparametric Regression and Generalized Linear Models*. London: Chapman Hall, 1994.

[6] J. A. Fessler, "Tomographic reconstruction using information-weighted spline smoothing," in *Information Processing in Medical Imaging* (H. H. Barrett and A. F. Gmitro, eds.), pp. 290–300, Berlin: Springer-Verlag, 1993.

[7] S. X. Pan and A. C. Kak, "A computational study of reconstruction algorithms for diffraction tomography: Interpolation versus filtered backpropagation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 1262–1275, 1983.

[8] H. Stark, J. W. Woods, I. Paul, and R. Hingorani, "Direct Fourier reconstruction in computer tomography," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 29, pp. 237–245, 1981.

[9] H. Stark, J. W. Woods, I. Paul, and R. Hingorani, "An investigation of computerized tomography by direct Fourier inversion and optimum interpolation," *IEEE Trans. Biomed. Eng.*, vol. 28, pp. 496–505, 1981.

[10] P. R. Smith, T. M. Peters, and R. H. T. Bates, "Image reconstruction from a finite number of projections," *J. Phys. A*, vol. 6, pp. 361–382, 1973.

[11] H. Stark, "Sampling theorems in polar coordinates," *J. Opt. Soc. Am.*, vol. 69, pp. 1519–1525, 1979.

[12] H. Stark and M. Wengrovitz, "Comments and corrections on the use of polar sampling theorems in CT," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 1329–1331, 1983.

[13] D. Fraser, "Interpolation by the FFT revisited—an experimental investigation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 665–675, 1989.

[14] T. J. Cavicchi, "DFT time-domain interpolation," *IEE Proceedings-F*, vol. 139, pp. 207–211, 1992.

[15] P. J. La Rivière and X. Pan, "Mathematical equivalence of zero-padding and circular sampling thorem interpolation with implications for direct Fourier image reconstruction," in *Proc. SPIE*, vol. 3338, pp. 1117–1126, 1998.

[16] E. O. Brigham, *The Fast Fourier Transform and Its Applications*. Englewood Cliffs: Prentice Hall, 1988.

[17] P. J. La Rivière and X. Pan, "Comparison of angular interpolation approaches in few-view tomography using statistical hypothesis testing," in *Proc. SPIE*, vol. 3661, 1999. (in press).

[18] L. A. Shepp and B. F. Logan, "The Fourier reconstruction of a head section," *IEEE Trans. Nucl. Sci.*, vol. 21, pp. 21–43, 1974.

[19] H. Spath, *One-Dimensional Spline Interpolation Algorithms*. Wellesley, MA: A. K. Peters, 1993.

[20] J. A. Rice, *Mathematical Statistics and Data Analysis*. Belmont, CA: Wadsworth, 1994.

[21] P. McCullagh and J. A. Nelder, *Generalized Linear Models*. London: Chapman Hall, 1983.

[22] N. P. Galatsanos and A. K. Katsaggelos, "Method for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Tr. Im. Proc.*, vol. 1, pp. 322–336, 1992.

[23] G. Wahba, *Spline Models for Observational Data*. Philadelphia: SIAM Press, 1990.

[24] F. O'Sullivan, B. S. Yandell, and W. J. Raynor, Jr., "Automatic smoothing of regression functions in generalized linear models," *J. Amer. Stat. Assoc.*, vol. 81, pp. 96–103, 1986.

[25] M. F. Hutchinson and F. R. de Hoog, "Smoothing noisy data with spline functions," *Numer. Math.*, vol. 47, pp. 99–106, 1985.

[26] E. V. Garcia, K. V. Train, J. Maddahi, F. Prigent, J. Friedman, J. Areeda, A. Waxman, and D. S. Berman, "Quantification of rotational thallium-201 myocardial tomography," *J. Nucl. Med.*, vol. 26, pp. 17–26, 1985.

Nonparametric regression sinogram smoothing using a roughness-penalized Poisson likelihood objective function

Patrick J. La Rivière, *Student Member, IEEE,* and Xiaochuan Pan,

Member, IEEE

May 24, 2000

Abstract

We develop and investigate an approach to tomographic image reconstruction in which nonparametric regression using a roughness-penalized Poisson likelihood objective function is used to smooth each projection independently prior to reconstruction by unapodized filtered backprojection (FBP). As an added generalization, the roughness penalty is expressed in terms of a monotonic transform, known as the link function, of the projections. The approach is compared to shift-invariant projection filtering through the use of a Hanning window as well as to a related nonparametric regression approach that makes use of an objective function based on weighted least squares (WLS) rather than the Poisson likelihood. The approach is found to lead to improvements in resolution-noise tradeoffs over the Hanning filter as well as over the WLS approach. We also investigate the resolution and noise effects of three different link functions: the identity, square root, and logarithm links. The choice of link function is found to influence the resolution uniformity and isotropy properties of the reconstructed images. In particular, in the case of an idealized imaging system with intrinsically uniform and istropic resolution, the choice of a square root link function yields the desirable outcome of essentially uniform and isotropic resolution in reconstructed images, with noise performance still superior to that of the Hanning filter as well as that of the WLS approach.

Keywords

Image reconstruction, sinogram smoothing, nonparametric regression, nonuniform resolution

I. INTRODUCTION

Fully iterative, statistical image reconstruction algorithms represent the most principled solution to the emission tomography image reconstruction problem, and steady improvements in computing power, the development of more efficient algorithms, and a deeper understanding of convergence and regularizing penalties have begun to win a place for these algorithms in the clinical setting. Nonetheless, perhaps because of its computational efficiency, straightforward filtered backprojection (FBP) using an apodized ramp filter is still widely used.

A significant drawback of FBP using a spatially invariant filter is that it treats all of the projection data equally, neglecting the data's inherently nonstationary statistics [1]. Conversely, one of the principal advantages of the fully iterative reconstruction methods is their explicit modeling of the projection data statistics. Fessler [1] has argued that much of this statistical benefit, with little of the computational cost, might be captured by principled one-dimensional (1D) smoothing of each projection followed by reconstruction using

unapodized FBP. In this vein, he proposed fitting to each projection a curve minimizing a roughness-penalized, weighted least squares (WLS) objective function in which the weights were inversely proportional to the data values (he also included a factor reflecting differences in detector efficiency and attenuation, which we disregard in this work in order to focus exclusively on statistical variability) [1]. This is a form of roughness-penalized nonparametric regression (NPR), in which the usual assumptions of classical linear regression are relaxed to allow for more general curves to be fit [2].

The WLS component of the objective function is optimal only for normally distributed data with unequal variances. Whereas randoms-corrected positron emission tomography (PET) data may be approximately normally distributed, unmodified PET and single-photon emission computed tomography (SPECT) data generally follow a Poisson distribution. For this reason, we recently introduced a roughness-penalized NPR projection-smoothing technique based on an explicit Poisson likelihood model for use in these situations [3]. In the present paper, we describe this approach in greater detail and present an important generalization involving the freedom to choose the so-called *link function* of the NPR model. As we shall see, different link functions induce slightly different resolution-noise tradeoffs in reconstructed images but more importantly lead to significantly different resolution uniformity and isotropy properties.

When images have uniform resolution, it is often characterized by a single, shift-invariant point-spread function (PSF) reflecting the response of the system, including the reconstruction algorithm, to an impulse input. When resolution is nonuniform, however, a generalization of the PSF concept is needed to reflect the shift-variant and possibly object-dependent nature of the resolution. Such a generalization is provided by the concept of a local impulse response (LIR) [4] function, which is the system response to impulses added to particular positions in particular objects. In this work, we use the term *nonuniform* resolution to mean that the breadth of the LIR function, and specifically of its half-maximum contour, varies as a function of position within the image and possibly also as a function of the object being imaged. We use the term *anisotropic* resolution to mean that the LIR functions at some or all image points are not circularly symmetric, which indicates that resolution is different in different directions. The presence of nonuniform and anisotropic

image resolution, while arguably desirable in some contexts, is often considered detrimental to image quality. Nonuniform resolution means, for instance, that identical lesions in different parts of the image would appear to have different sizes; anisotropic resolution means, for instance, that circular structures in the object would have an elliptical or otherwise distorted appearance in the image.

The factors affecting the uniformity or isotropy of reconstructed tomographic images can be divided into two broad categories: those intrinsic to the imaging system and those related to the reconstruction algorithm. In order to focus on the factors pertaining to reconstruction algorithms, the imaging system modeled in this work will be assumed to have intrinsically uniform and isotropic resolution properties. Among real emission tomography systems, PET comes closest to satisfying this assumption, especially for objects confined to the central region of the field of view. SPECT systems generally have intrinsically depth-dependent spatial resolution. Even in the face of projection data with intrinsically uniform and isotropic resolution, different reconstruction algorithms have the ability to produce images with different resolution properties. In a continuous formulation, simple FBP, with its shift-invariant filtering, naturally leads to images with uniform and isotropic resolution,[1] but, in general, statistically based approaches produce images with nonuniform, anisotropic resolution. Indeed, Fessler and Rogers have shown both analytically and empirically [4] that in emission tomography, fully iterative, penalized WLS and Poisson likelihood reconstruction approaches with a quadratic penalty lead to a particularly undesirable kind of nonuniform resolution in which resolution is better in low-count than in high-count areas. Moreover, these approaches also lead to anisotropic resolution that can distort structures in reconstructed images. Fessler and his collaborators have shown that it is possible to design object-dependent penalties for fully iterative approaches that lead to essentially uniform [4] and isotropic [7] resolution, but the process is somewhat involved and adds another computationally intensive step to the reconstruction process.

[1]The usual discretization of the FBP algorithm [5, Ch. 8] can introduce some slight resolution nonuniformities into reconstructed images. Guédon and Bizais [6] have proposed an alternative implementation, involving a filtering kernel that depends on the projection angle, that retains the shift-invariance of resolution in the discrete reconstructed image. Nonetheless, because the deviation from resolution uniformity in standard FBP is small compared to that arising when making use of some of the projection smoothing approaches under consideration, we have used the standard discretization of FBP throughout.

We will demonstrate in this work the ability to control resolution uniformity and isotropy properties through the choice of the link function in the nonparametric regression sinogram smoothing method, and we will highlight the ability to achieve essentially uniform and isotropic resolution through the choice of a square-root link function. We will also compare the resolution-noise tradeoffs achievable with the proposed Poisson likelihood-based objective function with those obtained by the use of two other sinogram smoothing approaches: simple Fourier-domain apodization windows and roughness-penalized nonparametric regression based on a WLS rather than a Poisson likelihood objective function. Statistical variability will be the principal physical deviation from ideal projection data that is considered. If the proposed approach is found to yield significant advantages over these alternative 1D approaches, it will be extended in future work to incorporate corrections for additional physical factors such as scatter, attenuation, and depth-dependent blurring, by the means discussed in Sec. II, and comparison then made to fully iterative approaches.

II. THEORY

A. Model and Assumptions

We begin with a projection model similar to that of Fessler [1], although with a few important differences with respect to the corrections applied to the measured data and the resulting statistical distribution of the data. Let $t(\xi, \eta)$ represent the two-dimensional activity distribution in the object being imaged expressed in a Cartesian coordinate system $\{\xi, \eta\}$ that is rotated by angle ϕ relative to a fixed $\{x, y\}$ coordinate system [8]. The Radon transform of this activity distribution for angle ϕ is given by

$$p_{\text{ideal}}^{(\phi)}(\xi) = \int_{-\infty}^{+\infty} t(\xi, \eta) d\eta, \tag{1}$$

where ϕ is known as the *projection angle* and ξ is known as the *projection distance*.

In general, this Radon transform is degraded by attenuation, blurring, and scatter. For example, in SPECT with an attenuation map $a(\xi, \eta)$, a possibly depth-dependent blurring function $k^{(\phi)}(\xi, \xi', \eta)$, and a scatter contribution $s^{(\phi)}(\xi)$ to each projection, we can express these degraded projections, which we denote simply $p^{(\phi)}(\xi)$, as

$$p^{(\phi)}(\xi) = \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} t(\xi', \eta) k^{(\phi)}(\xi, \xi', \eta) d\xi' \right] A(\xi, \eta, \phi) d\eta + s^{(\phi)}(\xi), \qquad (2)$$

where

$$A(\xi, \eta, \phi) = \exp \left\{ - \int_{\eta}^{\infty} a(\xi, \eta') d\eta' \right\}. \qquad (3)$$

Similar equations hold for PET, although in that case the attenuation factor does not depend on η, a detector efficiency factor should be added, and an additional additive term $r^{(\phi)}(\xi)$ reflecting the contribution of random coincides should be added to Eq. 2.

In practice, one measures projections at a finite number of angles ϕ_j, $j = 1, \ldots, M$, equally spaced over π or 2π, and each projection is sampled[2] at a finite number of projection distances or *bins* ξ_i, $i = 1, \ldots, N$. We denote the measurement at projection angle ϕ_j and projection bin ξ_i as $y_i^{(j)}$, and because the measurements exhibit statistical variability, they will be regarded as realizations of random variables $Y_i^{(j)}$ having means $\mu_i^{(j)} \equiv E\left\{Y_i^{(j)}\right\} = p^{(\phi_j)}(\xi_i)$, where $E\left\{\cdot\right\}$ denotes the expectation. The distribution of these random variables will be assumed to be Poisson, although the strategy discussed below can accommodate other exponential-family distributions as well [9].[3] As a point of notation, in what follows, the vector $\mathbf{y}^{(j)} \equiv \left(y_1^{(j)}, \ldots, y_N^{(j)}\right)^T$ will denote the set of measurements at projection angle ϕ_j and the vector $\mathbf{y} \equiv \left(\left[\mathbf{y}^{(1)}\right]^T, \ldots, \left[\mathbf{y}^{(M)}\right]^T\right)^T$ will denote the concatenation of these projection vectors into a single vector of measured data.

The goal in tomographic reconstruction is to estimate the activity distribution t, and more specifically some discrete approximation of it, from the measurements $\mathbf{y}$. Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_P)^T$ denote the discrete approximation of t that is to be estimated, where the θ_m, $m = 1, \ldots, P$, are lexicographically ordered pixel values, with P being the total number of pixels. Much work has been performed on the "direct" estimation of $\boldsymbol{\theta}$ from $\mathbf{y}$

---

[2]Note that we are modeling the measurements as point samples of $p^{(\phi_j)}(\xi)$. This will simplify the notation and analysis to follow, although a relatively straightforward extension of the formalism would allow the measurements to be modeled more accurately as integrals over the finite width of the detector elements, as in [1].

[3]The Poisson assumption is excellent for raw SPECT and PET data. PET measurements that have been precorrected for random coincidences do not follow a strictly Poisson distribution. Fessler's WLS objective function [1] may be the best choice in that situation, although it is possible that the corrected PET data may be well approximated by another exponential-family distribution, in which case the proposed penalized likelihood approach would be most appropriate. The resolution of that question is beyond the scope of the present paper.

by use of fully iterative algorithms such as maximum-likelihood estimation-maximization [10], maximum a posteriori estimation [11], penalized WLS [12], and the like. These approaches can in principle model and compensate for the attenuation, blurring, scattering, and statistical variability discussed above. The results of these approaches are promising, but the computational burden is high and many questions regarding regularization and convergence remain unresolved. An alternative, "indirect" path to the estimation of $\boldsymbol{\theta}$ is to estimate $p_{\text{ideal}}^{(\phi_j)}(\xi)$ from $\mathbf{y}^{(j)}$ as best as one can independently at each angle ϕ_j, and then to reconstruct the image by means of an algorithm such as FBP [1].

Direct estimation of $p_{\text{ideal}}^{(\phi_j)}(\xi)$, compensating for noise, scatter, blurring, and attenuation is an ambitious proposition, perhaps best handled by applying a sequence of corrections to the projections. In this paper, we will focus on the first step of such a sequence, the statistical problem of fitting smooth curves to the noisy measurements $\mathbf{y}^{(j)}$ in order to estimate $p^{(\phi_j)}(\xi)$. Future extensions of the approach will focus on correcting the estimates of $p^{(\phi_j)}(\xi)$ for scatter, blurring, and attenuation in order to estimate $p_{\text{ideal}}^{(\phi_j)}(\xi)$. For example, in SPECT, once estimates $\widehat{p}^{(\phi_j)}(\xi)$ of the $p^{(\phi_j)}(\xi)$ are available, scatter could be corrected for by subtracting estimates $\widehat{s}^{(\phi_j)}(\xi)$ of the scatter functions, obtained either from scatter windows or from analytic scatter estimation techniques of the sort developed by Frey [13]. In the case of SPECT-like blurring that is depth-dependent but shift-invariant in any plane parallel to the detector plane (i.e., when $k^{(\phi)}(\xi, \xi', \eta)$ in Eq. 2 can be written $k^{(\phi)}(\xi - \xi', \eta)$), the techniques of Metz and Pan [14, 15] could then be applied to the projection data to estimate the ideal projections. In fact, the approach proposed in this work could be used to estimate the $\widehat{s}^{(\phi_j)}(\xi)$ from counts collected in the scatter window or to estimate transmission sinograms for use in constructing attenuation maps when transmission data has been acquired as well. Still, these extensions are beyond the scope of the present paper, where we will focus exclusively on describing a method for the estimation of the $p^{(\phi_j)}(\xi)$ from the noisy data $\mathbf{y}^{(j)}$ and on assessing the noise and resolution properties of images reconstructed by use of this method.

Although it is not usually couched in terms of estimation theory, the most common way to estimate $p^{(\phi_j)}(\xi)$ from $\mathbf{y}^{(j)}$ in tomography is through the use of Fourier-domain apodization windows. In this approach, the discrete Fourier transform (DFT) of each

vector $\mathbf{y}^{(j)}$ is computed and the result multiplied by an apodization window such as a Hanning or Butterworth filter (this is normally performed in concert with the multiplication of the data by the ramp filter needed for accurate reconstruction) prior to taking an inverse DFT. The window suppresses the highest frequency components, presumed to contain more noise than signal, and this process is equivalent to performing a convolution of the data with a shift-invariant smoothing kernel. However, this kind of shift-invariant filtering is optimal only for stationary data, whereas the noise in tomographic projections is often quite nonstationary. A more sophisticated approach can be found in the work of Benali *et al.* [16] and Pélégrini *et al.* [17,18], who apply to emission tomography projection data a transformation of the form $\sqrt{y_i^{(j)} + 3/8}$, which changes the Poisson variables into asymptotic Gaussian variables with variance 0.25. They then smooth the transformed curve by use of a statistical technique known as the fixed-effect model. The similarities and differences between this approach and ours will be discussed in greater detail in Sec. V-B. In short, though, unlike the use of apodization windows and the approach just mentioned, the approach we propose explicitly models the Poisson nature of the projection data statistics. The curves $p^{(\phi_j)}(\xi)$ are estimated from $\mathbf{y}^{(j)}$ by maximizing objective functions that are derived using a statistical framework known as nonparametric regression with a generalized linear model (GLM) [2], which we will now describe.

B. Nonparametric Regression with a Generalized Linear Model

B.1 Statistical Model

Regression analysis with a single explanatory variable considers the problem of fitting a curve to a set of data pairs (y_i, ξ_i), $i = 1, \ldots, N$, where the y_i are the measured values of the quantity of interest and the ξ_i the corresponding values of the explanatory variable.[4] The variation in the y_i is assumed to have two components: a systematic component captured by a vector of predictors ψ_i that depends on the ξ_i, and a random component specifying the distribution of random variables Y_i, of which the y_i are realizations, given ψ_i [2]. In classical linear regression, for example, the systematic component is assumed to be of the

---

[4]In the sections describing the nonparametric regression approach in general, we use the same variables ξ and y as in the sections dealing specifically with the tomographic reconstruction problem, although for notational simplicity we omit the superscript $^{(j)}$ that labels the projection angle in those sections.

form $\psi_i = a\xi_i + b$, and the Y_i are assumed to be normally distributed about the ψ_i, i.e., $Y_i \sim N(\psi_i, \sigma^2)$, where σ^2 denotes the variance. In general, maximum-likelihood methods are used to estimate regression curves, and in the case of classical linear regression, the maximum-likelihood estimates of a and b are easily shown [19] to be those that minimize the sum of squared differences $\sum_{i=1}^{N}(y_i - a\xi_i - b)^2$.

Nonparametric regression using a GLM relaxes both of the assumptions of classical linear regression. First, it relaxes the assumption that the data are normally distributed in favor of the broader class of exponential distributions, which have probability densities of the form

$$f_{Y_i}(y_i \mid \psi_i, \varphi) = \exp\left(\frac{y_i \psi_i - b(\psi_i)}{\varphi} + c(y_i, \varphi)\right), \tag{4}$$

where ψ_i is the so-called natural parameter of the exponential family and φ is a scale parameter [9]. This family comprises many well-known distributions, each corresponding to a different choice of the functions b and c. Of particular interest to emission tomography is the choice $b(\psi_i) = e^{\psi_i}$, $\varphi = 1$, $c(y_i, \varphi) = -\log(y_i!)$, which corresponds to a Poisson distribution with mean $\mu_i = e^{\psi_i}$.

Second, nonparametric regression using a GLM eliminates the assumption that the predictors ψ_i depend on the explanatory variable in a simple, known parametric way, relating them instead to an unknown, continuous function of the explanatory variable, $q(\xi)$ [2]. This kind of nonparametric model is clearly desirable in the smoothing of emission tomography measurements, where the explanatory variable, ξ, is a spatial coordinate, and there is no reason to assume that the projections have any particular functional form. The most straightforward way to proceed is to connect this regression function $q(\xi)$ directly to the means of the random variables (which are themselves related in a known way to the predictors ψ_i) through the relation $\mu_i = q(\xi_i)$. Then the estimate of the regression function $q(\xi)$ would be a direct estimate of the curve of interest, which we denote $p(\xi)$. However, it is worthwhile to consider a more general formulation in which the regression function is connected to some monotonic transform $g(\cdot)$ of the means of the random variables

$$g(\mu_i) = q(\xi_i), \tag{5}$$

where $g(\cdot)$ is known in the NPR literature as a *link function*. Defining $h(\cdot) = g^{-1}(\cdot)$, this

can be rewritten

$$\mu_i = h\left[q(\xi_i)\right]. \tag{6}$$

It is important to appreciate that the task at hand is the estimation of continuous regression function $q(\xi)$. The link function $g(\cdot)$ is specified as part of the model. Possible choices of the link function include the identity link, $g(\mu_i) = \mu_i$, the logarithm link, $g(\mu_i) = \ln(\mu_i)$, and the square root link, $g(\mu_i) = \sqrt{\mu_i}$.

For a given link function, the regression function $q(\xi)$ is, in general, a transformation of the actual curve of interest. For example, in the emission tomography application, our interest is in fitting a curve $\widehat{p}^{(\phi_j)}(\xi)$ to the measurements $\mathbf{y}^{(j)}$ at each projection angle ϕ_j. With the choice, for example, of the logarithm link $g(\mu_i) = \ln(\mu_i)$, the estimate $\widehat{q}^{(\phi_j)}(\xi)$ yielded by this framework is an estimate of the logarithm of the estimate $\widehat{p}^{(\phi_j)}(\xi)$ of interest. Only for the identity link is the estimate $\widehat{q}^{(\phi_j)}(\xi)$ a direct estimate of $\widehat{p}^{(\phi_j)}(\xi)$.

This generalization has important implications for resolution uniformity and isotropy properties. As discussed below, the regression curves $\widehat{q}^{(\phi_j)}(\xi)$ will be estimated by maximizing a penalized-likelihood objective function. Working in terms of a monotonic transform of the curve of interest does not affect the likelihood component of the objective function but it does cause the smoothing penalty to be enforced on this transform of the curve of interest. The measurement variance "perceived" by the smoothing penalty is affected by the transform and influences the resulting resolution properties of the reconstructed images. This point will be discussed in greater detail in Sec. V.

All of the results of this work will be derived in terms of a general link function $g(\cdot)$, assumed to be monotonic so that $h(\cdot)$ exists. We also assume that all needed derivatives of $g(\cdot)$ and $h(\cdot)$ exist. We will then examine the effect of a few particular choices of the link function on the resolution and noise properties in reconstructed images.

B.2 Curve Fitting

The generalization of parametric regression to nonparametric regression would seem to complicate the problem immensely, turning a relatively straightforward finite-dimensional estimation problem into an intractable infinite-dimensional one. As in classical linear regression, it might seem wise to fit the curve by use of maximum likelihood methods,

but in the nonparametric case, *any* curve $q(\xi)$ for which $h\left[q(\xi_i)\right] = y_i$, that is, any curve that interpolates the data no matter how it behaves between the samples, maximizes the likelihood [2]. In practice the problem is made tractable by adding the further constraint that the estimated curve $q(\xi)$ be smooth and enforcing this constraint by penalizing the log likelihood with a term of the form $-\frac{1}{2}\lambda \int \left[q''(\xi)\right]^2 d\xi$, where $q''(\xi)$ is the second derivative of $q(\xi)$ and λ is the smoothing parameter [2]. If the unpenalized likelihood depends on $q(\xi)$ only through its values $q(\xi_i)$ at the measured points ξ_i, $i = 1, \ldots, N$, it can be shown that the maximizer of the penalized likelihood is necessarily a natural cubic spline [2]. These are curves that are piecewise cubic polynomials between each pair of measured points and that satisfy the so-called "natural" boundary conditions of being first-order (linear) polynomials to the left of the first measured point and to the right of the last measured point. Such curves, which are continuous up to and including the first derivative, can be specified by a finite number of coefficients.[5]

The log-likelihood for exponential densities is given by

$$\ln L(\boldsymbol{\psi}, \varphi, \mathbf{y}) = \sum_{i=1}^{N} \left(\frac{y_i \psi_i - b(\psi_i)}{\varphi} + c(y_i, \varphi) \right). \tag{7}$$

We now define $q_i \equiv q(\xi_i)$ and a vector $\mathbf{q} = (q_1, \ldots, q_N)^T$. Substituting the functions b and c appropriate for a Poisson density, using the fact, stated above, that $\mu_i = e^{\psi_i}$ for the Poisson distribution, and substituting for μ_i using Eq. 6, we have

$$\ln L(\mathbf{q}, \mathbf{y}) = \sum_{i=1}^{N} \left\{ y_i \ln \left[h(q_i) \right] - h(q_i) \right\}, \tag{8}$$

where we have dropped terms independent of q_i.

For natural cubic splines, the penalty term takes on a simple form: $-\frac{1}{2}\lambda \int \left[q''(\xi)\right]^2 d\xi = -\frac{1}{2}\lambda \mathbf{q}^T K \mathbf{q}$, where K is an $N \times N$ matrix determined by the spacing of the measurement points ξ_i [2] and defined in Appendix A. Because the natural cubic spline interpolating any specified set of points $q(\xi_i)$ is unique [2], finding $\hat{q}(\xi)$ maximizing Eq. 8 penalized by

---

[5]In this sense, roughness-penalized nonparametric regression does yield a parametrized curve, but given that there are at least as many parameters as observations, the spirit is rather different from that of standard parametric techniques.

$-\frac{1}{2}\lambda \int [q''(\xi)]^2\, d\xi$ is thus tantamount to finding $\widehat{\mathbf{q}}$ maximizing the objective function

$$\Phi(\mathbf{q}, \mathbf{y}) = \sum_{i=1}^{N} \{y_i \ln [h(q_i)] - h(q_i)\} - \frac{1}{2}\lambda \mathbf{q}^T K \mathbf{q}. \tag{9}$$

The $\widehat{\mathbf{q}}$ maximizing this objective function can be found by use of Fisher scoring [9], which entails an iterative update of the form

$$\mathbf{q}^{new} = \mathbf{q} + \left(E\left\{ -\frac{\partial^2 \Phi}{\partial \mathbf{q} \partial \mathbf{q}^T} \right\} \right)^{-1} \frac{\partial \Phi}{\partial \mathbf{q}}, \tag{10}$$

which can be shown to yield

$$\mathbf{q}^{new} = (W + \lambda K)^{-1} W \mathbf{z}, \tag{11}$$

where W is a diagonal matrix with diagonal elements

$$W_{ii} = \frac{[h'(q_i)]^2}{h(q_i)}, \tag{12}$$

and $\mathbf{z}$ is given by

$$z_i = q_i + \frac{1}{h'(q_i)} [y_i - h(q_i)]. \tag{13}$$

The process is initialized by setting the $q_i = g(y_i)$. For the identity and square root link functions, care must be taken to avoid dividing by zero when elements of $\mathbf{q}$ grow small. In these cases we set $q_i = \max\{q_i, \epsilon\}$ for some small $\epsilon > 0$. While each iteration might appear to be computationally intense, requiring the computation of the inverse of the $N \times N$ matrix $(W + \lambda K)$, in practice many of the matrices involved have band properties that can be exploited to keep the computation to $\vartheta(N)$ [2]. Iteration continues until the change in $h(\mathbf{q})$ from one iteration to the next falls below some small, specified threshold.

B.3 Image Reconstruction

Returning to the tomographic reconstruction problem, the strategy is thus to apply this nonparametric regression curve-fitting to each projection independently. For each projection angle ϕ_j, we take the measurements $\mathbf{y}^{(j)}$ and find $\widehat{\mathbf{q}}^{(j)}$ maximizing Eq. 9 by means of the algorithm just described. The elements of the vector $\widehat{\mathbf{q}}^{(j)}$ uniquely specify a continuous spline $\widehat{q}^{(\phi_j)}(\xi)$, and we could easily evaluate this spline at any points ξ desired. The estimate of the corresponding smoothed projection values would then be given

by $\widehat{p}^{(\phi_j)}(\xi) = h\left[\widehat{q}^{(\phi_j)}(\xi)\right]$. However, because reconstruction will be performed by use of FBP, we generally require estimates of $\widehat{p}^{(\phi_j)}(\xi)$ only at the same positions ξ_i at which the original measurements were obtained; there is no need ever to construct explicitly the spline specified by $\widehat{\mathbf{q}}^{(j)}$. The elements $\widehat{q}_i^{(j)}$, which are samples of the smooth curve $\widehat{q}^{(\phi_j)}(\xi)$, are all that are required, and from them we estimate the samples $\widehat{p}_i^{(j)} = h\left(\widehat{q}_i^{(j)}\right)$ of the smooth curve $\widehat{p}^{(\phi_j)}(\xi)$ needed for reconstruction. This will be expressed in vector notation as $\widehat{\mathbf{p}}^{(j)} = h\left(\widehat{\mathbf{q}}^{(j)}\right)$, where here and throughout, the function h of a vector is assumed to operate element by element. Reconstruction can then proceed by FBP, using an unapodized ramp filter because the projections have already been smoothed [5].

The overall reconstruction process can thus be summarized by the equation

$$\widehat{\boldsymbol{\theta}}(\mathbf{y}) = \sum_{j=1}^{M} \mathcal{B}^{(j)} \left\{ h\left[\arg\max_{\mathbf{q}^{(j)}} \Phi(\mathbf{q}^{(j)}, \mathbf{y}^{(j)})\right] \right\}, \tag{14}$$

where $\mathcal{B}^{(j)}$ is a matrix operator representing the ramp filtering and backprojection of the discrete samples $h\left[\arg\max_{\mathbf{q}^{(j)}} \Phi(\mathbf{q}^{(j)}, \mathbf{y}^{(j)})\right]$ at projection angle ϕ_j.

One advantage of fully iterative algorithms that cannot be reproduced exactly by this approach is the ability to enforce nonnegativity constraints on reconstructed pixel values. FBP simply does not enforce this constraint. At least two of the link functions do, however, have the desirable property of enforcing nonnegativity in the smoothed projections. The square root link $g(\mu_i) = \sqrt{\mu_i}$, and the logarithm link $g(\mu_i) = \ln(\mu_i)$, have inverses $h(\mu_i) = \mu_i^2$ and $h(\mu_i) = \exp(\mu_i)$, respectively, that automatically guarantee nonnegativity in the smoothed projections. Moreover, for the identity link, the guard against allowing values of q_i to reach zero exactly also discourages the q_i from crossing below zero.

B.4 An Alternative: Roughness-Penalized WLS

For comparison, we will also examine the application of a roughness-penalized WLS objective function similar to the one proposed by Fessler [1]. In this approach, each fit curve $\widehat{p}(\xi)$ is the minimizer of the objective function

$$\Phi_{WLS}(p(\xi), \mathbf{y}) = \sum_{i=1}^{N} w_i \left[y_i - p(\xi_i)\right]^2 + \lambda \int \left[p''(\xi)\right]^2 d\xi. \tag{15}$$

The objective function is written in terms of $p(\xi)$ rather than $q(\xi)$ because we will not consider link functions other than the identity in this case; the smoothed projections will be

estimated directly. As with the Poisson likelihood objective function, it would of course be possible to incorporate different link functions into the WLS objective function and thus to engender a range of resolution uniformity properties. However, because the two approaches make different assumptions about the data statistics, our true interest is in comparing the *noise* properties of images reconstructed by use of the WLS approach to those of images reconstructed by use of the Poisson likelihood approach, and this can be achieved most directly when using the identity link for both. The w_i here are weights reflecting the relative certainty of each measurement and we set them to be $w_i = 1/\max\{y_i, \epsilon\}$, for some small $\epsilon > 0$.

As with Eq. 9, Eq. 15 is minimized by a natural cubic spline, and thus the penalty term again takes on a simple form, $\int [p''(\xi)]^2 d\xi = \mathbf{p}^T K \mathbf{p}$, where $\mathbf{p} = (p_1, \ldots, p_N)^T$ with $p_i \equiv p(\xi_i)$. Equation 15 can thus be rewritten in vector form

$$\Phi_{WLS}(\mathbf{p}, \mathbf{y}) = (\mathbf{y} - \mathbf{p})^T W (\mathbf{y} - \mathbf{p}) + \lambda \mathbf{p}^T K \mathbf{p}, \tag{16}$$

where now W is a diagonal matrix with diagonal elements w_i. Unlike Eq. 9, which required an iterative method to find the maximum, the minimum of Eq. 16 can be solved for in closed form. It can be shown [2] that the $\hat{\mathbf{p}}$ minimizing Eq. 16 is

$$\hat{\mathbf{p}} = (W + \lambda K)^{-1} W \mathbf{y}. \tag{17}$$

This equation has a very similar structure to Eq. 11. In fact, for the Poisson likelihood objective function with the identity link, Eq. 11 produces an iteratively reweighted least squares algorithm that differs from Eq. 17 only in that the weights used in the W matrix are determined iteratively from the y_i rather than estimated directly from them.

Despite the existence of a closed-form solution for the WLS objective, we will still express the image-reconstruction estimator in terms of an extremum of the objective function, in order to maintain a notation parallel to that used in the Poisson likelihood case:

$$\hat{\boldsymbol{\theta}}_{WLS}(\mathbf{y}) = \sum_{j=1}^{M} \mathcal{B}^{(j)} \left\{ \arg \min_{\mathbf{p}^{(j)}} \Phi_{WLS}(\mathbf{p}^{(j)}, \mathbf{y}^{(j)}) \right\}. \tag{18}$$

III. Theoretical Resolution and Noise Analysis

A. Local Impulse Response

A.1 Penalized Poisson Likelihood

As mentioned in the introduction, when characterizing imaging systems and algorithms possessing nonuniform resolution properties, the local impulse response function concept provides a useful generalization of the point-spread function concept. The LIR is the system response to impulses added to particular positions in particular objects and is defined as

$$l^m(\boldsymbol{\theta}) = \frac{\partial}{\partial \theta_m} \boldsymbol{\mu}(\boldsymbol{\theta}), \quad m = 1, \ldots, P, \tag{19}$$

where $\boldsymbol{\mu}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \left[\widehat{\boldsymbol{\theta}}(\mathbf{Y}) \right]$ is the mean of the estimator.

In [4], Fessler and Rogers discuss an unbiased empirical estimator of LIR that allows for the evaluation of LIR for any pixel m from a set of R images estimated from R noisy realizations $\left\{ \mathbf{y}_{(r)} \right\}_{r=1}^{R}$ of the projections of the object of interest. It is given by

$$\widehat{l^m}(\boldsymbol{\theta}) = \frac{1}{R-1} \sum_{r=1}^{R} \left(\widehat{\boldsymbol{\theta}}(\mathbf{y}_{(r)}) - \widehat{\boldsymbol{\mu}}(\boldsymbol{\theta}) \right) \frac{\partial \log f_{\mathbf{Y}}(\mathbf{y}_{(r)}; \boldsymbol{\theta})}{\partial \theta_m}, \tag{20}$$

where $\widehat{\boldsymbol{\mu}}(\boldsymbol{\theta}) = \frac{1}{R} \sum_{r=1}^{R} \widehat{\boldsymbol{\theta}}(\mathbf{y}_{(r)})$ is the sample mean of the reconstructions and $f_{\mathbf{Y}}(\mathbf{y}; \boldsymbol{\theta})$ is the statistical distribution of the measurements.

Because R may need to be large to obtain accurate estimates of LIR (we use $R = 250$ in the studies discussed below), an analytic expression for LIR would clearly be of great benefit. For many well-behaved estimators, a locally linear approximation [4] allows the LIR to be expressed in terms of the estimator's action on a set of noiseless data:

$$l^m(\boldsymbol{\theta}) = \frac{\partial}{\partial \theta_m} \widehat{\boldsymbol{\theta}}(\overline{\mathbf{Y}}(\boldsymbol{\theta})), \quad m = 1, \ldots, P, \tag{21}$$

where $\overline{\mathbf{Y}}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}[\mathbf{Y}]$. As a basis for empirical evaluations, this is clearly a great improvement over Eq. 20 because it allows the estimator LIR for a pixel of interest to be computed from two reconstructions: one from noiseless projections of the object and one from noiseless projections of the object plus an impulse at the pixel of interest [4].

In addition to providing a less intensive recipe for empirical assessment of LIR, Eq. 21 can also be evaluated analytically. Using the expression for $\widehat{\boldsymbol{\theta}}(\mathbf{y})$ in Eq. 14, we have

$$\mathbf{l}^m(\boldsymbol{\theta}) = \sum_{j=1}^{M} \mathcal{B}^{(j)} \left\{ \frac{\partial}{\partial \theta_m} h \left[\arg \max_{\mathbf{q}^{(j)}} \Phi(\mathbf{q}^{(j)}, \overline{\mathbf{Y}}^{(j)}) \right] \right\}. \tag{22}$$

Evaluating Eq. 22 entails computing the partial derivatives of a function of the implicitly defined estimators $\widehat{\mathbf{q}}^{(j)} = \arg \max_{\mathbf{q}^{(j)}} \Phi(\mathbf{q}^{(j)}, \mathbf{y}^{(j)})$. We do not have an explicit form for the estimators, but fortunately, the partial derivatives can be computed solely from the objective functions themselves using the implicit function theorem and the chain rule [4]. Specifically, as discussed in Appendix B, the expression $\widehat{\mathbf{q}}^{(j)} = \arg \max_{\mathbf{q}^{(j)}} \Phi(\mathbf{q}^{(j)}, \mathbf{y}^{(j)})$ implicitly defines the estimator $\widehat{\mathbf{q}}^{(j)}$ as a function of the measurement $\mathbf{y}^{(j)}$:

$$\widehat{\mathbf{q}}^{(j)} = f^{(j)}(\mathbf{y}^{(j)}) = \left[f_1^{(j)}(\mathbf{y}^{(j)}), \ldots, f_N^{(j)}(\mathbf{y}^{(j)}) \right], \tag{23}$$

which allows Eq. 22 to be evaluated, as shown in Appendix B. The result found there is

$$\mathbf{l}^m(\boldsymbol{\theta}) = \sum_{j=1}^{M} \mathcal{B}^{(j)} \left\{ A \left[h(\breve{\mathbf{q}}^{(j)}), \overline{\mathbf{Y}}^{(j)} \right] \frac{\partial}{\partial \theta_m} \overline{\mathbf{Y}}^{(j)}(\boldsymbol{\theta}) \right\}, \tag{24}$$

where $A \left[h(\breve{\mathbf{q}}^{(j)}), \overline{\mathbf{Y}}^{(j)} \right]$ is a matrix defined in Appendix B and $\breve{\mathbf{q}}^{(j)} = \arg \max_{\mathbf{q}^{(j)}} \Phi(\mathbf{q}^{(j)}, \overline{\mathbf{Y}}^{(j)})$ is the result of applying the nonparametric regression estimator to the noisefree projection $\overline{\mathbf{Y}}^{(j)}$. Table I evaluates $A \left[h(\breve{\mathbf{q}}^{(j)}), \overline{\mathbf{Y}}^{(j)} \right]$ for three different link functions: the identity link, for which $g(\mu_i) = \mu_i$ and $h(\mu_i) = \mu_i$; the logarithm link, for which $g(\mu_i) = \ln(\mu_i)$ and $h(\mu_i) = e^{\mu_i}$; and the square root link, for which $g(\mu_i) = \sqrt{\mu_i}$ and $h(\mu_i) = \mu_i^2$. In the table, $D[x_i]$ denotes a diagonal matrix with diagonal elements x_i.

Equation 24 is easily evaluated in practice. Evaluating the braces for all projection angles yields a set of vectors that can be assembled into an array with the dimensions of a sinogram, which can be reconstructed by FBP to obtain the LIR for pixel m. Within the braces, $\frac{\partial}{\partial \theta_m} \overline{\mathbf{Y}}^{(j)}(\boldsymbol{\theta})$ is the projection at angle ϕ_j of an impulse at pixel m, the $\overline{Y}_i^{(j)}$ correspond to the noiseless projections of the object of interest at that angle, and the $\breve{q}_i^{(j)}$ are the values obtained by applying the penalized Poisson likelihood estimator for the link function in question to the noiseless projection. It should be pointed out that only the $\frac{\partial}{\partial \theta_m} \overline{\mathbf{Y}}^{(j)}(\boldsymbol{\theta})$ need be reevaluated to obtain the LIR for different pixels. However, the

whole process must be repeated to evaluate the LIR for different values of the smoothing parameter λ.

The validity of the locally linear approximation that licenses Eq. 21 and thus ultimately Eq. 22 will be evaluated below by comparing results obtained using this approximation with fully empirical results.

A.2 Penalized WLS

Under the locally linear approximation, the same analysis can be applied to the WLS estimator with an identity link of Eq. 18. In this case,

$$l^m_{WLS}(\boldsymbol{\theta}) = \sum_{j=1}^{M} \mathcal{B}^{(j)} \left\{ \frac{\partial}{\partial\theta_m} \left[\arg\min_{\mathbf{p}^{(j)}} \Phi(\mathbf{p}^{(j)}, \overline{\mathbf{Y}}^{(j)}) \right] \right\}. \tag{25}$$

Once again, the derivatives can be evaluated by invoking the implicit function theorem and the chain rule. The details are given in Appendix C, which leads to the conclusion that

$$l^m_{WLS}(\boldsymbol{\theta}) = \sum_{j=1}^{M} \mathcal{B}^{(j)} \left\{ A\left[\check{\mathbf{p}}^{(j)}, \overline{\mathbf{Y}}^{(j)}\right] \frac{\partial}{\partial\theta_m} \overline{\mathbf{Y}}^{(j)}(\boldsymbol{\theta}) \right\}, \tag{26}$$

where

$$A\left[\check{\mathbf{p}}^{(j)}, \overline{\mathbf{Y}}^{(j)}\right] = \left\{ D\left[\frac{1}{\overline{Y_i}^{(j)}}\right] + \lambda K \right\}^{-1} D\left[\frac{\check{p}_i^{(j)}}{\overline{Y_i}^{(j)^2}}\right]. \tag{27}$$

This expression can be evaluated in the same way as Eq. 24.

B. Variance

B.1 Penalized Poisson Likelihood

As with the evaluation of resolution properties, empirical evaluation of estimator variance can be extremely computationally intensive, but once again the locally linear approximation can be used to derive an analytic expression. We begin with the definition of the estimator covariance

$$\mathrm{Cov}\left\{\widehat{\boldsymbol{\theta}}\right\} = E\left\{ \left[\widehat{\boldsymbol{\theta}} - E\left\{\widehat{\boldsymbol{\theta}}\right\}\right] \left[\widehat{\boldsymbol{\theta}} - E\left\{\widehat{\boldsymbol{\theta}}\right\}\right]^T \right\}. \tag{28}$$

Substituting Eq. 14 for $\widehat{\boldsymbol{\theta}}$ and expressing $\widehat{\mathbf{q}}^{(j)} = \arg\max_{\mathbf{q}^{(j)}} \Phi(\mathbf{q}^{(j)}, \mathbf{y}^{(j)})$ in terms of the implicit function of Eq. 23, this can easily be shown to yield

$$\mathrm{Cov}\left\{\widehat{\boldsymbol{\theta}}\right\} = \sum_{j=1}^{M} \mathcal{B}^{(j)} \mathrm{Cov}\left\{h\left[f^{(j)}(\mathbf{Y}^{(j)})\right]\right\} \left[\mathcal{B}^{(j)}\right]^{T}. \tag{29}$$

Under the locally linear approximation [20], the central factor, the covariance of the estimator for each projection, is given by

$$\mathrm{Cov}\left\{h\left[f^{(j)}(\mathbf{Y}^{(j)})\right]\right\} = \nabla h\left[f^{(j)}(\overline{\mathbf{Y}}^{(j)})\right] \mathrm{Cov}\left\{\mathbf{Y}^{(j)}\right\} \left[\nabla h\left[f^{(j)}(\overline{\mathbf{Y}}^{(j)})\right]\right]^{T}. \tag{30}$$

The central factor here, $\mathrm{Cov}\left\{\mathbf{Y}^{(j)}\right\}$, is simply given by $D\left[\overline{\mathbf{Y}}_i^{(j)}\right]$ for uncorrelated Poisson noise, and the other factor, $\nabla h\left[f^{(j)}(\overline{\mathbf{Y}}^{(j)})\right]$ is given by

$$\nabla h\left[f^{(j)}(\overline{\mathbf{Y}}^{(j)})\right] = D\left[h'(f_i^{(j)}(\overline{\mathbf{Y}}^{(j)}))\right] \nabla f^{(j)}(\overline{\mathbf{Y}}^{(j)}), \tag{31}$$

where $\nabla f^{(j)}(\overline{\mathbf{Y}}^{(j)})$ is evaluated in Eq. 37 of Appendix B.

Evaluating the entire $P \times P$ covariance matrix $\mathrm{Cov}\left\{\widehat{\boldsymbol{\theta}}\right\}$ would be quite computationally intense, but this is often not necessary. In the present work, we will be concerned only with the variance of the estimator $\widehat{\boldsymbol{\theta}}$ for particular pixels m. Defining e^m to be the $1 \times N$ unit vector with a 1 in the mth element, we can write

$$
\begin{aligned}
\mathrm{Var}\left\{\widehat{\theta}_m\right\} &= (e^m)^T \mathrm{Cov}\left\{\widehat{\boldsymbol{\theta}}\right\} e^m \\
&= \sum_{j=1}^{M} (e^m)^T \mathcal{B}^{(j)} \mathrm{Cov}\left\{h\left[f^{(j)}(\mathbf{Y}^{(j)})\right]\right\} \mathcal{B}^{(j)T} e^m.
\end{aligned} \tag{32}
$$

This expression can be evaluated as follows. The $1 \times N$ vector $(e^m)^T \mathcal{B}^{(j)}$ represents the mth row of the filtration and backprojection matrix $\mathcal{B}^{(j)}$ for projection angle ϕ_j. Each element of this vector thus represents the contribution of a particular projection bin to image pixel m. The ith element of the vector is obtained by filtering and backprojecting a projection at angle ϕ_j consisting simply of a unit impulse in the ith projection bin and picking up the value in pixel m of the reconstruction matrix. The central factor, $\mathrm{Cov}\left\{h\left[f^{(j)}(\mathbf{Y}^{(j)})\right]\right\}$, can be computed using Eqs. 30 and 31 as well as Eq. 37 of Appendix B.

Once again, the validity of the locally linear approximation that underlies this approach will be evaluated below by comparing results obtained using this approximation with fully empirical results.

B.2 Penalized WLS

For the PWLS projection smoothing, the analytic derivation of noise properties is essentially identical. All the results of the previous section hold with the function h taken to be the identity function and Eq. 42 of Appendix C used to evaluate $\nabla \left[f^{(j)}(\overline{\mathbf{Y}}^{(j)}) \right]$ when needed.

IV. RESULTS

A. *Local Impulse Response*

A.1 Validation of Locally Linear Approximation

In order to validate the locally linear approximation that is the basis of Eqs. 21 and 24 for the penalized Poisson likelihood approach and Eq. 26 for the penalized WLS approach, we computed the unbiased empirical estimator of LIR given in Eq. 20 and compared it to the analytic predictions. We did so for both the penalized WLS approach with the identity link and the penalized Poisson likelihood approach with three choices of link function: identity, logarithm, and square root.

Specifically, we took as our object $\boldsymbol{\theta}$ the phantom shown in Fig. 1, having relative emission intensities 1, 2, and 3 in the cold disk (left), background, and hot disk (right), respectively. This phantom, which is essentially the same as the one used in [4], will allow us to test the dependence of LIR on count levels. The phantom was specified on a 128×128 grid and we generated projections of 128 bins and 128 angles, using strip integrals the same width as the bins, which in turn were the same size as the pixels. The system thus has intrinsically uniform and isotropic resolution. We did not model the effects of attenuation or scatter. The mean number of counts per bin was approximately 6–10. We generated 250 Poisson realizations of the projections and performed reconstructions using each of the four methods under consideration for values of the smoothing parameter chosen to yield approximately the same average image resolution in all four cases. We evaluated the LIR at three different pixels: one in the center of the cold disk, one in the center of the background ellipse, and one in the center of the hot disk. We also computed the analytic predictions of LIR at these positions using Eqs. 24 and 26.

Figure 2 displays horizontal profiles through the LIR for the estimators corresponding

to these four reconstruction methods. The profiles through the LIRs for the cold disk, background ellipse, and hot disk are concatenated left to right in this figure. The triangles are for the empirical estimator of Eq. 20 and the solid line for the analytic evaluation of Eqs. 24 (penalized Poisson likelihood) and 26 (penalized WLS).

The agreement between the empirical results and the analytic prediction is clearly excellent, and justifies our use of the locally linear approximation—at least for resolution analysis—in what follows.

A.2 Resolution Nonuniformity

One fact that is clear upon inspection of Fig. 2 is that the smoothing approaches lead to images with nonuniform, object-dependent resolution, and that the choice of link function has a profound effect on this property. The LIR for penalized Poisson likelihood with an identity link can be seen to be wider in the high-count region than in the low-count region, indicating that resolution is better in low-count than in high-count areas. The logarithm link leads to the opposite behavior: the LIR is wider in the low-count region than in the high-count region, indicating that resolution is better in high-count than in low-count areas. Finally, for the square root link, the LIRs seem to be of more or less constant width, indicating that resolution is essentially uniform and object independent. While they are not shown, LIR profiles for images reconstructed after smoothing the projections with a Hanning filter have a similar appearance to those for the penalized Poisson likelihood estimator with a square root link: constant width for all three profiles. Not surprisingly, since the link function is the chief determinant of resolution uniformity properties, the penalized WLS estimator with an identity link exhibits the same behavior as the penalized Poisson likelihood estimator with an identity link. For clarity, we will thus omit the WLS results from the remainder of the resolution uniformity studies in order to focus on link-function effects; we will return to the WLS estimator only when we consider noise properties.

These resolution uniformity trends were verified by a further study: LIRs at these three locations were computed by evaluating the analytic expression for LIR for the penalized Poisson likelihood with these three link functions and a range of values of the smoothing parameter, as well as empirically for the Hanning filter with a range of filter cutoffs. The

horizontal and vertical full-width half-maxima (FWHM) of each LIR were then computed, as well as an average FWHM, which we define as the average of the horizontal and vertical FWHM.[6] In Fig. 3, we plot the ratio of the average FWHM of the hot-disk LIR to that of the cold-disk LIR. For each estimator, these ratios are plotted versus the average FWHM of the center (background ellipse) LIR so that they correspond to more or less similar degrees of smoothing for the various methods. We can see clearly that the nonuniformity trends observed in the profiles of Fig. 2 are typical for a wide range of values of the smoothing parameter, though not surprisingly they diminish as the smoothing parameter grows small and the influence of the penalty term in the objective function becomes negligible.

While we have been assessing the resolution *uniformity* properties of the estimators by comparing the average FWHM of the LIRs in high- and low-count regions, it is equally important to consider the *isotropy* of resolution at a given point in the image as captured by the symmetry or asymmetry of the LIR for that pixel. That some of the methods lead to asymmetric LIRs is evident from Fig. 4, which displays contours at levels 25, 50, 75, and 99% of peak value for each of the LIRs whose horizontal profiles were shown in Fig. 2. As in the case of resolution uniformity, the identity and logarithm links lead to essentially opposite behavior, and the square root link yields essentially symmetric LIRs. While they are not shown, the contours for LIRs after projection smoothing with a simple Hanning filter are also essentially circular.

Once again, these trends were verified by a further study. The analytically computed LIRs used to generate the hot/cold LIR ratios of Fig. 3 were this time used to generate ratios of horizontal FWHM to vertical FWHM for the LIR corresponding to the background ellipse. The results are plotted in Fig. 5, again versus the average FWHM of background ellipse LIR so that they correspond to more or less similar degrees of smoothing for the various methods.

The results presented thus far have been for three locations in the phantom, but they are typical of results observed at other positions, including those at boundaries between regions in the phantom. To confirm this assertion quantitatively, we calculated the following global measure of image uniformity and isotropy. We calculated the LIR at 60 different positions

---

[6]This definition of average FWHM is meaningful here because the FWHM contours at these three image locations are all approximately elliptical with major and minor axes aligned horizontally and vertically.

in the phantom when reconstructing by FBP after projection smoothing by use of the penalized Poisson likelihood estimator with the three link functions as well as by use of the Hanning filter. We selected values of the smoothing parameters and the Hanning filter cutoff to yield approximately equal values of the global average radius (3.6 cm) of the 50% contour for all four methods. For each method we calculated this global average, as well as the mean absolute deviation of the radii of the 50% contour from this average. Small values of the deviation indicate that resolution is essentially uniform and isotropic. The logarithm and identity links led to a mean absolute deviation of 0.13 and 0.11 pixels, respectively, while the square root link and Hanning filtration each led to a mean absolute deviation of 0.02 pixels.

The source of resolution nonuniformity and anisotropy, the dependence on the choice of link function, and the consequences of these properties will be discussed further in Sec. V.

B. Variance

B.1 Validation of Locally Linear Approximation

Although the locally linear approximation underlying the analytic expressions for LIR was shown by the results of Fig. 2 to be justified in that context, that does not necessarily mean that the same approximation is justified for the derivation of analytic variance expressions. It is possible that the noise properties of the reconstructed images could depend more strongly on the non-locally linear components of the estimator than do the resolution properties, and this is indeed found to be the case. We computed empirical noise levels by generating 800 Poisson realizations of the projections of the phantom in Fig. 1 and reconstructing by each of the methods under consideration for a range of values of the smoothing parameter. We then computed the sample standard deviation, as well its 95% confidence interval, at three points in the image: the center of the hot disk, the center of the background ellipse, and the center of the cold disk. We also computed the analytic prediction of standard deviation at those three points for all the methods and smoothing parameters using Eq. 32.

For all the methods, the analytic predictions generally lay within the 95% confidence interval of the empirical estimate of standard deviation for all values of the smoothing

parameter for the point at the center of the cold disk and the point at the center of the phantom. However, for the point at the center of the hot disk, the analytic predictions for all the methods generally lay below the 95% confidence interval of the empirical standard deviation for larger values of the smoothing parameter. A typical such result is shown in Fig. 6, where the analytic predictions and empirical estimates (with 95% confidence intervals) of standard deviation for the Poisson likelihood estimator with an identity link for the point at the center of the hot disk are plotted versus the analytic prediction of average LIR FWHM at that point, giving the plots the form of resolution-noise curves. It is not surprising that the agreement between the analytic predictions and empirical results always holds for small values of the smoothing parameter, because the approaches are actually linear in the limit as the smoothing parameter approaches zero. Nonetheless, it is the larger values of the smoothing parameter that are likely to be of interest in practical situations, and the occasional failure of the locally linear assumption in this regime suggests that we should not rely upon it to characterize the noise properties of the approaches.

C. Resolution-Noise Tradeoff

We have seen above that the different link functions under consideration induce different, count-dependent resolution uniformity and isotropy properties. It is of course also important to consider the resolution-noise tradeoffs induced by the various methods. As we have seen in the previous sections, the analytic expression for LIR agrees well with empirical results while the analytic expression for variance does not, so we have generated resolution-noise curves in which the resolution properties are determined using the analytic expression for LIR while the variance results are generated from empirical simulations involving 800 realizations, as discussed in the previous section.

In order to give an overall sense of comparative resolution-noise tradeoffs for the WLS estimator with an identity link, the Poisson likelihood estimator with the identity, logarithm, and square root links, and the Hanning filter, we have plotted them together in two separate figures: one for the point at the center of the cold disk (Fig. 7(a)) and one for the point at the center of the hot disk (Fig. 7(b)). The Hanning filter curve was generated using empirical calculations of both resolution and noise properties. It does not

extend below approximately 2 pixels average FWHM because it is not possible to achieve such low degrees of smoothing with a Hanning filter. As it is, the range of the curve calls attention to the region of the bias-variance curves that is truly of practical interest—the region involving moderate amounts of smoothing of the degree encountered when making use of an apodization window in FBP. Close-ups of these regions are depicted in Figs. 7(c) and (d). The fact that all the objective-function-based approaches behave nearly identically in the upper-left, low-smoothing region is explained by the fact that as the smoothing parameter approaches zero, the approaches all yield solutions that approach the unconstrained maximum likelihood solution of simply interpolating the measurements.

One must take some care in making comparisons between the curves on this plot because they are for estimators with different resolution isotropy properties. Considering only the average FWHM of the LIR tends to handicap estimators with isotropic resolution properties, because it is reasonable to believe that for a given average resolution there exists some anisotropic smoothing that can outperform the isotropic smoothing in terms of noise properties [7]. This said, it is encouraging to see that the resolution-noise performance of the Poisson likelihood estimator with a square root link, which we have seen to have quite isotropic resolution properties, generally performs better than the WLS estimator with the identity link and the Poisson likelihood estimator with a logarithm link and that it performs comparably to the Poisson likelihood estimator with an identity link, all of which we have seen to have anisotropic resolution properties.

It is more meaningful to compare estimators with matched resolution isotropy properties, and we adopt this approach to resolve the question of whether there is indeed any advantage in adopting the estimator based on the Poisson likelihood over that based on WLS or the Hanning filter. In particular, in Fig. 8(a) we plot resolution-noise curves with 95% confidence intervals for the points at the center of hot and cold disks for the Poisson likelihood estimator and for the WLS estimator, both with an identity link to ensure matched resolution properties. For the moderate to high levels of smoothing (≥ 3 pixels average FWHM) of practical interest, the curves for the Poisson likelihood estimator lies below and to the left of the curves for the WLS estimator and the confidence intervals do not overlap, indicating superior performance for the Poisson likelihood estimator. In

Fig. 8(b), we plot resolution-noise curves with 95% confidence intervals for the points at the center of hot and cold disks for the Poisson likelihood estimator with a square root link and for the Hanning filter. Both have isotropic resolution properties, as we have seen. We observe that over much of the range of interest, the curves for the Poisson likelihood estimator lie below and to the left of the curves for the Hanning filter and that the confidence intervals for the two different approaches do not overlap, again indicating superior performance for the Poisson likelihood estimator.

V. Discussion

A. The Link Function

The forms of the expressions for the matrix $A\left[h(\breve{\mathbf{q}}^{(j)}), \overline{\mathbf{Y}}^{(j)}\right]$ given in Table I for the Poisson likelihood estimators with different link functions offer some insight into the nonuniform, count-dependent image resolution that arises when applying these estimators. Consider, as in [4], the simplified situation where the measurement noise has constant variance σ^2. In this case, the matrix A would become:

$$
\begin{aligned}
\text{IDENTITY}: &\quad \left\{I + \sigma^2 \lambda K\right\}^{-1} \\
\text{LOG}: &\quad \left\{I + \frac{\lambda}{\sigma^2} K\right\}^{-1} \\
\text{SQUARE ROOT}: &\quad \left\{I + \frac{\lambda}{4} K\right\}^{-1}.
\end{aligned}
\tag{33}
$$

We see that for the identity link, the measurements with variance σ^2 induce an effective smoothing parameter $\sigma^2 \lambda$. Thus higher variance leads to more smoothing in this case. Similar behavior carries over to the more complicated case of nonconstant Poisson variance, where regions of higher variance (higher counts) in the projections are smoothed more than regions of lower variance (lower counts). Because regions of high counts in the projections are generally associated with regions of high counts in the image, the result is reconstructed images whose high-count regions are smoothed more heavily than are its low-count regions. Applying the logarithm link to projections with constant variance σ^2 leads to an effective smoothing parameter $\frac{\lambda}{\sigma^2}$, and thus the situation is the opposite of that just described: higher variance leads to less smoothing in this case and thus we would expect, in the

case of nonconstant, Poisson variance, reconstructed images whose low-count regions are smoothed more heavily than are its high-count regions. Finally, applying the square root link to projections with constant variance σ^2 leads to an effective smoothing parameter $\frac{\lambda}{4}$, which is independent of the variance σ^2. Thus projections are smoothed more or less equivalently regardless of count level, and the result, in the case of nonconstant Poisson variance, is reconstructed images with more or less uniform resolution.

The source of these properties becomes apparent when we consider more clearly the role the link function plays in projection smoothing. In developing the curve-fitting objective function, we effectively began with the log of the Poisson likelihood function,

$$\ln L(\boldsymbol{\mu}, \mathbf{y}) = \sum_{i=1}^{n} \left\{ y_i \ln \mu_i - \mu_i \right\}, \tag{34}$$

where terms independent of μ_i, the means of the random variables being measured, have been dropped. Introduction of the link function g means that one works to estimate not μ_i directly, but rather a transform $q(\xi_i) = g(\mu_i)$ of μ_i. For any monotonic link function g, unpenalized maximum likelihood estimation of the $q(\xi_i)$ would yield the same estimates $\mu_i = h(q(\xi_i)) = y_i$; thus, the introduction of a link function does not affect the log-likelihood component of the objective function. However, by expressing the roughness penalty $-\frac{1}{2}\lambda \int q''(\xi)^2 d\xi$ in terms of the curve $q(\xi)$ rather than the curve of interest $p(\xi) = h(q(\xi))$, we cause the roughness penalty to be enforced on the *transform* $q(\xi) = g(p(\xi))$ of the curve of interest. The variance "perceived" by the roughness penalty in these transformed curves is influenced by the choice of link function. As can readily be shown [19], if a random variable y with mean μ has variance $\sigma^2(\mu)$, then a transform $g(y)$ of the variable has variance $\approx \sigma^2(\mu) \left[g'(\mu)\right]^2$. For a Poisson random variable with $\sigma^2(\mu) = \mu$, the logarithm transform leads to a variance $1/\mu$ whereas a square root transform leads to an essentially constant variance equal to $1/4$. The identity transform naturally does not affect the variance at all; it remains equal to μ. These "perceived" variances agree perfectly with the effective smoothing parameters seen in Eq. 33 for these choices of link function. In fact, the logarithm link function was the first non-identity link function we investigated, and it was in coming to this understanding of the source of its nonuniform resolution behavior that we were lead to identify the square root link as likely to lead to essentially uniform resolution properties.

B. A Related Spline-Based Projection Smoothing Approach

Having described in some detail our nonparametric regression sinogram smoothing approach, it is perhaps worth noting some of its similarities and differences with the approach of Benali *et al.* [16] and Pélégrini *et al.* [17, 18] mentioned in Sec. I. In their approach, after applying the transformation $\sqrt{y_i^{(j)} + 3/8}$ to the projection data, they fit to each projection a cubic spline minimizing an unweighted, roughness-penalized least squares objective function. This is tantamount to constraining the smooth function underlying the transformed projection data to lie in a Sobolev space $H^{(2)}$, which is the space of functions whose derivatives up to order 1 are absolutely continuous and whose 2nd derivative is square integrable. Given the set of spline-smoothed projections, they then solve another optimization problem to find the L-dimensional subspace of $H^{(2)}$ containing the L principal components of the set of smoothed projections (where L is some freely chosen integer less than N, the number of measurements per projection). The smoothed projections are then projected onto this subspace and the inverse of the transformation $\sqrt{y_i^{(j)} + 3/8}$ applied to the result.

In our approach, the addition of the penalty term based on the integral of the squared second derivative to the Poisson likelihood or WLS objective functions in Eqs. 9 and 15 is also tantamount to constraining the solutions to lie in a Sobolev space $H^{(2)}$, although the differences in form between our objective functions and theirs mean that our solutions will not, in general, lie at the same point in the space as the solutions to their first optimization problem. Their use of a square root transform may seem to be similar to our use of a square-root link function, but there is an important difference. The square root transform allows them to *approximate* the Poisson-distributed random variables as normally distributed variables with constant variance, and thus to adopt an unweighted least squares objective function, whereas our approach, while incorporating the square root link to modify resolution properties, still retains the exact Poisson likelihood in the objective function. The significance of this difference for noise performance would be a worthy topic of future study. Finally, their second optimization, involving the reduction of the dimensionality of the Sobolev space inhabited by the smoothed projections has no counterpart in our model. It might be fruitful to explore the effect of this additional smoothing on

noise-resolution tradeoffs and particularly on resolution uniformity properties.

Benali *et al.* [16] and Pélégrini *et al.* [17,18] also investigate the application of an alternative implementation of FBP called spline-filtered backprojection (SFBP) that could potentially be incorporated into future extensions of our own approach. The appeal of SFBP is that if the smoothed projection data is a discrete version of projections belonging to a Sobolev space $H^{(m)}$, for integer m, then the approach guarantees that the reconstructed image will be a discrete version of a continuous activity distribution that belongs to a space $H^{(m-\frac{1}{2})}$. The approach consists essentially in oversampling the smoothed projection measurements using interpolating splines of degree $2m-1$, computing the discrete version of a spline-based, projection-angle-dependent reconstruction filter and convolving it with the oversampled projections, and finally backprojection the convolved projections using conventional algorithms. In guaranteeing that the regularity properties of the projections are carried into the reconstructed image, the approach also avoids the small image-space shift-variance introduced by the usual discretization of FBP.

VI. CONCLUSIONS

We have presented and investigated an approach to tomographic image reconstruction in which nonparametric regression using a roughness-penalized Poisson likelihood objective function is used to smooth each projection independently prior to reconstruction by un-apodized filtered backprojection (FBP). As an added generalization, the roughness penalty was expressed in terms of a monotonic transform, known as the link function, of the projections. The approach was found to have the desirable ability to produce images with essentially uniform and isotropic resolution properties when incorporating the square root link function. Of course, simple shift-invariant projection filtering also produces images with uniform and isotropic resolution, but the proposed sinogram smoothing approach was found to yield resolution-noise tradeoffs that were superior to those obtained when making use of shift-invariant filtering. The proposed approach was also found to outperform a second sinogram smoothing approach, based on a weighted least squares rather than Poisson likelihood objective function, when applied to genuinely Poisson-distributed data. In future work, the proposed approach will be extended to incorporate correction and compensation for a wider range of physical factors and then compared to fully iterative,

penalized-likelihood algorithms. We expect that the proposed sinogram smoothing and processing approach may provide a viable, computationally efficient alternative to these fully iterative approaches, particularly for relatively simple imaging tasks such as SPECT scintimammography [21].

APPENDIX A

This section defines the matrix K that arises in Eqs. 9 and 16, among others. K is defined in terms of two band matrices, S and R. Let $\Delta_i = \xi_{i+1} - \xi_i$, for $i = 1, \dots, N - 1$. Then S is the $N \times (N - 2)$ matrix with entries s_{ij}, for $i = 1, \dots, N$ and $j = 1, \dots, N - 2$, given by

$$s_{jj} = \Delta_j^{-1}, \ s_{j+1,j} = -\Delta_j^{-1} - \Delta_{j+1}^{-1}, \ s_{j+2,j} = \Delta_{j+1}^{-1}$$

for $j = 1, \dots, N - 2$, and $s_{ij} = 0$ for $i - j \geq 3$ or $i - j \leq -1$.

R is a symmetric $(N - 2) \times (N - 2)$ matrix with elements r_{ij} for $i = 1, \dots, N - 2$ and $j = 1, \dots, N - 2$, given by

$$
\begin{aligned}
r_{ii} &= \frac{1}{3}(\Delta_{i-1} + \Delta_i), \quad i = 1, \dots, N - 2, \\
r_{i,i+1} &= r_{i+1,i} = \frac{1}{6}\Delta_i, \quad i = 1, \dots, N - 3,
\end{aligned}
$$

and $r_{ij} = 0$ for $|i - j| \geq 2$. We can now give the definition of K:

$$K = SR^{-1}S^T.$$

This definition is always meaningful because it can be shown that R is strictly positive definite and thus R^{-1} necessarily exists. Further details are given in [2, p. 13], where, it should be noted, in defining K the authors use a non-standard matrix element numbering convention that simplifies their later derivations but that has not been used here.

APPENDIX B

Evaluating Eq. 22 entails computing the partial derivatives of a function of the implicitly defined estimators $\widehat{\mathbf{q}}^{(j)} = \arg\max_{\mathbf{q}^{(j)}} \Phi(\mathbf{q}^{(j)}, \mathbf{y}^{(j)})$. This can be accomplished by invoking the implicit function theorem [22] and the chain rule. The implicit function theorem holds that if $\Phi(\mathbf{q}^{(j)}, \mathbf{y}^{(j)})$ has a unique global maximum $\widehat{\mathbf{q}}^{(j)}$ for any $\mathbf{y}^{(j)}$, and that the partial

derivatives of $\Phi(\mathbf{q}^{(j)}, \mathbf{y}^{(j)})$ with respect to $\mathbf{q}^{(j)}$ are zero at $\widehat{\mathbf{q}}^{(j)}$, then there exists an implicit function $\widehat{\mathbf{q}}^{(j)} = f^{(j)}(\mathbf{y}^{(j)}) = \left[f_1^{(j)}(\mathbf{y}^{(j)}), \dots, f_N^{(j)}(\mathbf{y}^{(j)}) \right]$ satisfying

$$\frac{\partial}{\partial q_i^{(j)}} \Phi(f^{(j)}(\mathbf{y}^{(j)}), \mathbf{y}^{(j)}) \doteq 0, \quad i = 1, \dots, N, \tag{35}$$

where the notation should be interpreted to mean that the derivative of $\Phi(\mathbf{q}^{(j)}, \mathbf{y}^{(j)})$ with respect to $q_i^{(j)}$ is evaluated at $\mathbf{q}^{(j)} = f^{(j)}(\mathbf{y}^{(j)})$.[7] Thus, by the chain rule

$$\begin{aligned}
\frac{\partial}{\partial \theta_m} h \left[\arg\max_{\mathbf{q}^{(j)}} \Phi(\mathbf{q}^{(j)}, \overline{\mathbf{Y}}^{(j)}) \right] &= \frac{\partial}{\partial \theta_m} h \left[f^{(j)}(\overline{\mathbf{Y}}^{(j)}) \right] \\
&= D \left[h'(\check{q}_i^{(j)}) \right] \nabla f^{(j)}(\overline{\mathbf{Y}}^{(j)}) \frac{\partial}{\partial \theta_m} \overline{\mathbf{Y}}(\boldsymbol{\theta}),
\end{aligned} \tag{36}$$

where the notation $D[x_i]$ denotes a diagonal matrix with diagonal entries $D_{ii} = x_i$, $\check{\mathbf{q}}^{(j)} = \arg\max_{\mathbf{q}^{(j)}} \Phi(\mathbf{q}^{(j)}, \overline{\mathbf{Y}}^{(j)})$, which is the result of applying the nonparametric regression estimator to the noisefree projection $\overline{\mathbf{Y}}^{(j)}$, and $\nabla = \left[\frac{\partial}{\partial Y_1^{(j)}}, \dots, \frac{\partial}{\partial Y_N^{(j)}} \right]$.

It is straightforward to show [4] that

$$\nabla f^{(j)}(\overline{\mathbf{Y}}^{(j)}) = \left[-\nabla_{20} \Phi(\check{\mathbf{q}}^{(j)}, \overline{\mathbf{Y}}^{(j)}) \right]^{-1} \nabla_{11} \Phi(\check{\mathbf{q}}^{(j)}, \overline{\mathbf{Y}}^{(j)}), \tag{37}$$

where $\left[\nabla_{20} \Phi(\mathbf{q}^{(j)}, \mathbf{y}^{(j)}) \right]_{kl} = \frac{\partial^2}{\partial q_k^{(j)} q_l^{(j)}} \Phi(\mathbf{q}^{(j)}, \mathbf{y}^{(j)})$, and $\left[\nabla_{11} \Phi(\mathbf{q}^{(j)}, \mathbf{y}^{(j)}) \right]_{kl} = \frac{\partial^2}{\partial q_k^{(j)} \partial y_l^{(j)}} \Phi(\mathbf{q}^{(j)}, \mathbf{y}^{(j)})$. Using the definition of $\Phi(\mathbf{q}, \mathbf{y})$ in Eq. 9, it can then be shown that

$$\nabla_{20} \Phi(\mathbf{q}^{(j)}, \mathbf{y}^{(j)}) = D \left[\left\{ \frac{h''(q_i^{(j)})}{h(q_i^{(j)})} - \frac{h'(q_i^{(j)})^2}{h(q_i^{(j)})^2} \right\} y_i^{(j)} - h''(q_i^{(j)}) \right] - \lambda K, \tag{38}$$

and

$$\nabla_{11} \Phi(\mathbf{q}^{(j)}, \mathbf{y}^{(j)}) = D \left[\frac{h'(q_i^{(j)})}{h(q_i^{(j)})} \right]. \tag{39}$$

[7]The theorem's conditions are satisfied by the objective functions that arise for the three link functions being considered: identity, square root, and logarithm. Specifically, the Hessian $\nabla_{20} \Phi(\mathbf{q}^{(j)}, \mathbf{y}^{(j)})$ for a general link function is given in Eq. 38. The elements of the diagonal matrix evaluate to $-\frac{y_i^{(j)}}{\left(q_i^{(j)}\right)^2}$ for the identity link, $-\frac{y_i^{(j)}}{\left(q_i^{(j)}\right)^2} - 2$ for the square root link, and $-\exp(q_i^{(j)})$ for the logarithm link. Except when $y_i^{(j)} = 0$ for the identity link, the elements are strictly negative and the diagonal matrix is thus negative definite. Because λK can be shown to be nonnegative definite [2], the Hessian is thus negative definite and the objective function concave. The seemingly problematic situation when $y_i^{(j)} = 0$ for the identity link can be handled simply by discarding such zero-information measurements, as in [1].

Combining all these results gives the final expression for LIR given in Eq. 24, where we have defined

$$A\left[h(\breve{\mathbf{q}}^{(j)}), \overline{\mathbf{Y}}^{(j)}\right] = D\left[h'(\breve{q}_i^{(j)})\right] \nabla f^{(j)}(\overline{\mathbf{Y}}^{(j)}).$$

APPENDIX C

Using the argument of Appendix B,[8] the partial derivative of Eq. 25 can be written

$$\frac{\partial}{\partial \theta_m}\left\{\arg\min_{\mathbf{p}^{(j)}} \Phi(\mathbf{p}^{(j)}, \overline{\mathbf{Y}}^{(j)})\right\} = \nabla f^{(j)}(\overline{\mathbf{Y}}^{(j)})\frac{\partial}{\partial \theta_m}\overline{\mathbf{Y}}(\boldsymbol{\theta}), \tag{40}$$

with $f^{(j)}(\mathbf{y}^{(j)})$ being an implicit function that can be reexpressed for the noiseless data as

$$\nabla f^{(j)}(\overline{\mathbf{Y}}^{(j)}) = \left[-\nabla_{20}\Phi(\breve{\mathbf{p}}^{(j)}, \overline{\mathbf{Y}}^{(j)})\right]^{-1} \nabla_{11}\Phi(\breve{\mathbf{p}}^{(j)}, \overline{\mathbf{Y}}^{(j)}), \tag{41}$$

where $\breve{\mathbf{p}}^{(j)} = \arg\min_{\mathbf{p}^{(j)}} \Phi(\mathbf{p}^{(j)}, \overline{\mathbf{Y}}^{(j)})$ is the result of applying the WLS estimator to the noiseless data. In calculating the needed derivatives, we find that

$$\begin{aligned}
\nabla_{20}\Phi(\mathbf{p}^{(j)}, \mathbf{y}^{(j)}) &= 2D\left[\frac{1}{y_i^{(j)}}\right] + 2\lambda K \\
\nabla_{11}\Phi(\mathbf{p}^{(j)}, \mathbf{y}^{(j)}) &= -2D\left[\frac{p_i^{(j)}}{\left(y_i^{(j)}\right)^2}\right].
\end{aligned} \tag{42}$$

REFERENCES

[1] J. A. Fessler, "Tomographic reconstruction using information-weighted spline smoothing," in *Information Processing in Medical Imaging*, H. H. Barrett and A. F. Gmitro, Eds., pp. 290–300. Springer-Verlag, Berlin, 1993.

[2] P. J. Green and B. W. Silverman, *Nonparametric Regression and Generalized Linear Models*, Chapman Hall, London, 1994.

[3] P. J. La Rivière and X. Pan, "Few-view tomography using roughness-penalized nonparametric regression and periodic spline interpolation," *IEEE Trans. Nucl. Sci.*, vol. 46, pp. 1121–1128, 1999.

[4] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs," *IEEE Trans. Im. Proc.*, vol. 5, pp. 1346–1358, 1996.

[5] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Academic Press, New York, 1982.

[6] J.-P. Guédon and Y. Bizais, "Bandlimited and Haar filtered back-projection reconstructions," *IEEE Trans. Med. Imag.*, vol. 13, pp. 430–440, 1994.

[8]Here the Hessian can be shown to be positive definite, indicating that the objective function is convex and thus that the implicit function theorem, expressed in terms of minima rather than maxima, holds for all $\mathbf{y}^{(j)}$. (Recall that the WLS objective is *minimized* whereas the Poisson likelihood objective is *maximized*).

[7] J. W. Stayman and J. A. Fessler, "Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction," *IEEE Trans. Med. Imag.*, to appear.

[8] S. Deans, *The Radon Transform and Some of Its Applications*, Wiley, New York, 1983.

[9] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, Chapman Hall, London, 1983.

[10] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, pp. 113–122, 1982.

[11] E. Levitan and G. T. Herman, "A maximum *a posteriori* probability expectation maximization algorithm for image reconstruction in emission tomography," *IEEE Trans. Med. Imag.*, vol. 6, pp. 185–192, 1988.

[12] J. A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 13, pp. 290–300, 1994.

[13] E. C. Frey, Z.-W. Zu, and B. M. W. Tsui, "A fast projector-backprojector pair modeling the asymmetric, spatially varying scatter response function for scatter compensation in spect imaging," *IEEE Trans. Nucl. Sci.*, vol. 40, pp. 1192–1197, 1993.

[14] C. E. Metz and X. Pan, "A unified analysis of exact methods of inverting the 2-D exponential Radon transform, with implications for noise control in SPECT," *IEEE Trans. Med. Imag.*, vol. 14, pp. 643–658, 1995.

[15] X. Pan and C. E. Metz, "Non-iterative methods and their noise characteristics in 2D SPECT image reconstruction," *IEEE Trans. Nucl. Sci.*, vol. 44, pp. 1388–1397, 1997.

[16] H. Benali, J.-P. Guédon, I. Buvat, M. Pélégrini, Y. Bizais, and R. Di Paola, "A statistical model for tomographic reconstruction methods using spline functions," in *Proc. SPIE*, 1994, vol. 2299, pp. 242–251.

[17] M. Pélégrini, H. Benali, I. Buvat, G. El Fakhri, and R. Di Paola, "Two-dimensional statistical model for regularized backprojection in SPECT," *Phys. Med. Biol.*, vol. 43, pp. 421–434, 1998.

[18] M. Pélégrini, I. Buvat, H. Benali, G. El Fakhri, P. Grangeat, and R. Di Paola, "A spline-regularized minimal residual algorithm for iterative attenuation correction in SPECT," *Phys. Med. Biol.*, vol. 44, pp. 2623–2642, 1999.

[19] J. A. Rice, *Mathematical Statistics and Data Analysis*, Wadsworth, Belmont, CA, 1994.

[20] J. A. Fessler and W. L. Rogers, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Trans. Im. Proc.*, vol. 5, pp. 493–506, 1996.

[21] P. J. La Rivière, X. Pan, and B. C. Penney, "Ideal-observer analysis of lesion detectability in planar, conventional SPECT, and dedicated SPECT scintimammography using effective multi-dimensional smoothing," *IEEE Trans. Nucl. Sci.*, vol. 45, pp. 1273–1279, 1998.

[22] S. I. Grossman, *Multivariable Calculus, Linear Algebra, and Differential Equations*, Harcourt Brace Jovanovich, San Diego, 1986.

Link	$A\left[h(\check{\mathbf{q}}^{(j)}), \overline{\mathbf{Y}}^{(j)} \right]$
Ident.	$\left\{ D\left[\dfrac{\overline{Y}_i^{(j)}}{\left(\check{q}_i^{(j)}\right)^2} \right] + \lambda K \right\}^{-1} D\left[\dfrac{1}{\check{q}_i^{(j)}} \right]$
Log.	$D\left[\exp(\check{q}_i^{(j)}) \right] \left\{ D\left[\exp(\check{q}_i^{(j)}) \right] + \lambda K \right\}^{-1}$
Sqr. Rt.	$D\left[\check{q}_i^{(j)} \right] \left\{ D\left[\dfrac{2\overline{Y}_i^{(j)}}{(\check{q}_i^{(j)})^2} + 2 \right] + \lambda K \right\}^{-1} D\left[\dfrac{4}{\check{q}_i^{(j)}} \right]$

TABLE I

$A\left[h(\check{\mathbf{q}}^{(j)}), \overline{\mathbf{Y}}^{(j)} \right]$ OF EQ. 24 FOR 3 DIFFERENT LINK FUNCTIONS.
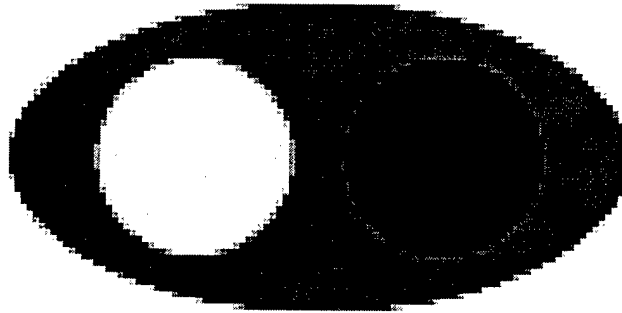
Fig. 1. Phantom used to examine resolution and noise properties of the estimators under consideration.

POI--Identity Link

POI--Logarithm Link

POI--Square Root Link

WLS--Identity Link
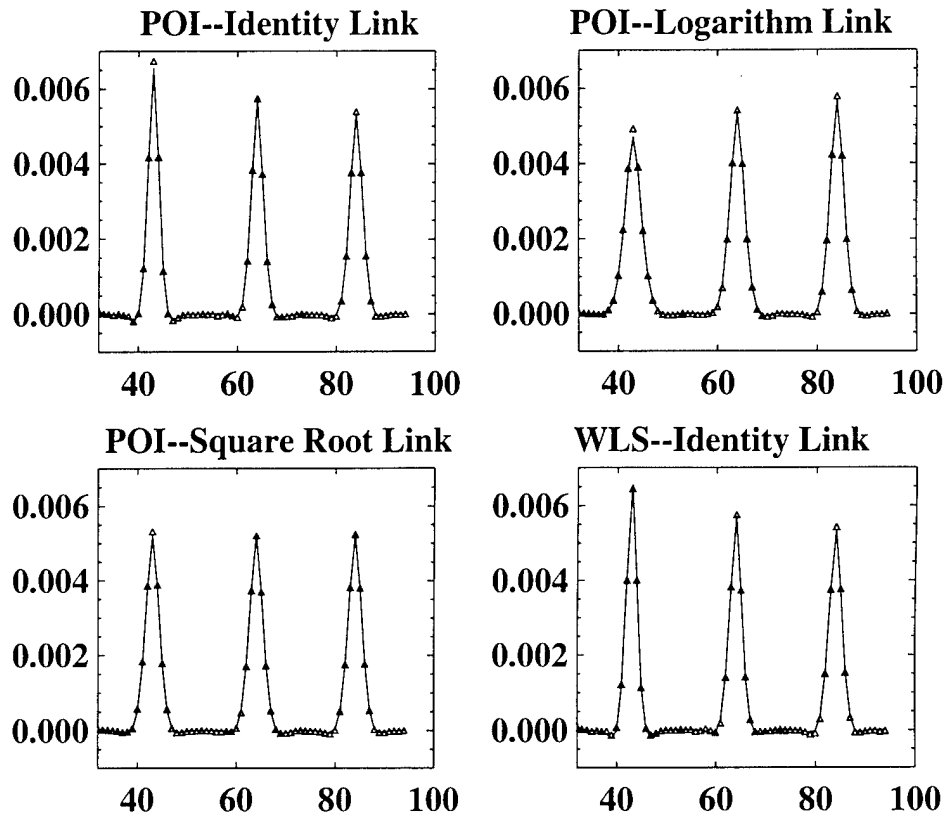
Fig. 2. Horizontal profiles through three LIR functions for four projection-smoothing methods under consideration applied to the phantom of Fig. 1. The profiles, which are concatenated left to right, are for the LIR functions corresponding to the points at the center of the cold disk, the center of the background ellipse, and the center of the hot disk, respectively. The triangles denote empirical results and the solid lines analytic predictions of LIR using Eqs. 24 and 26. The different link functions are seen to give rise to different resolution uniformity properties.
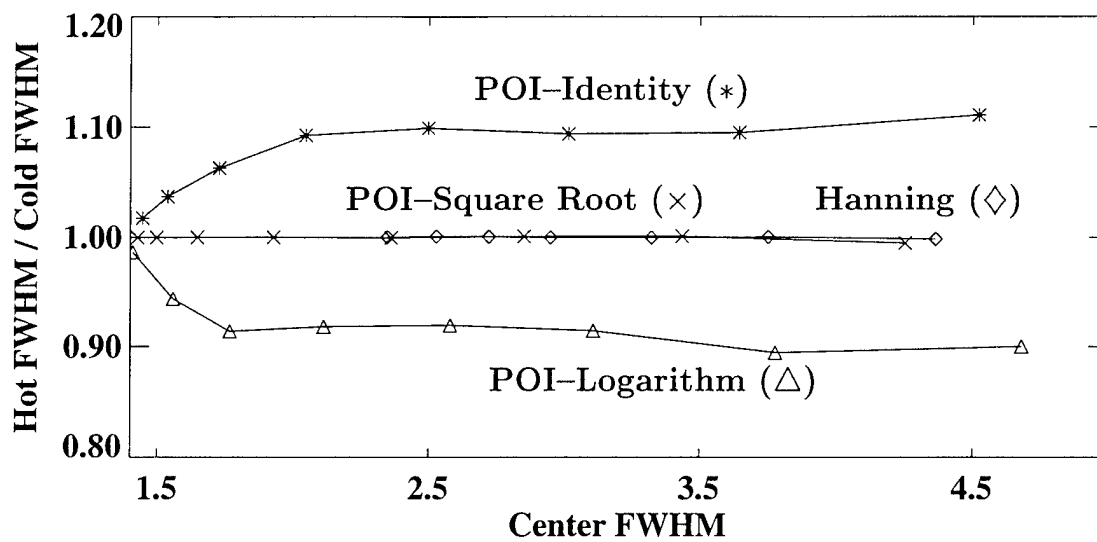
Fig. 3. Ratio of the average LIR function FWHM for the hot disk to that for the cold disk, plotted versus average FWHM for the center of the background ellipse. This figure confirms the nonuniform resolution properties observed in Fig. 2.
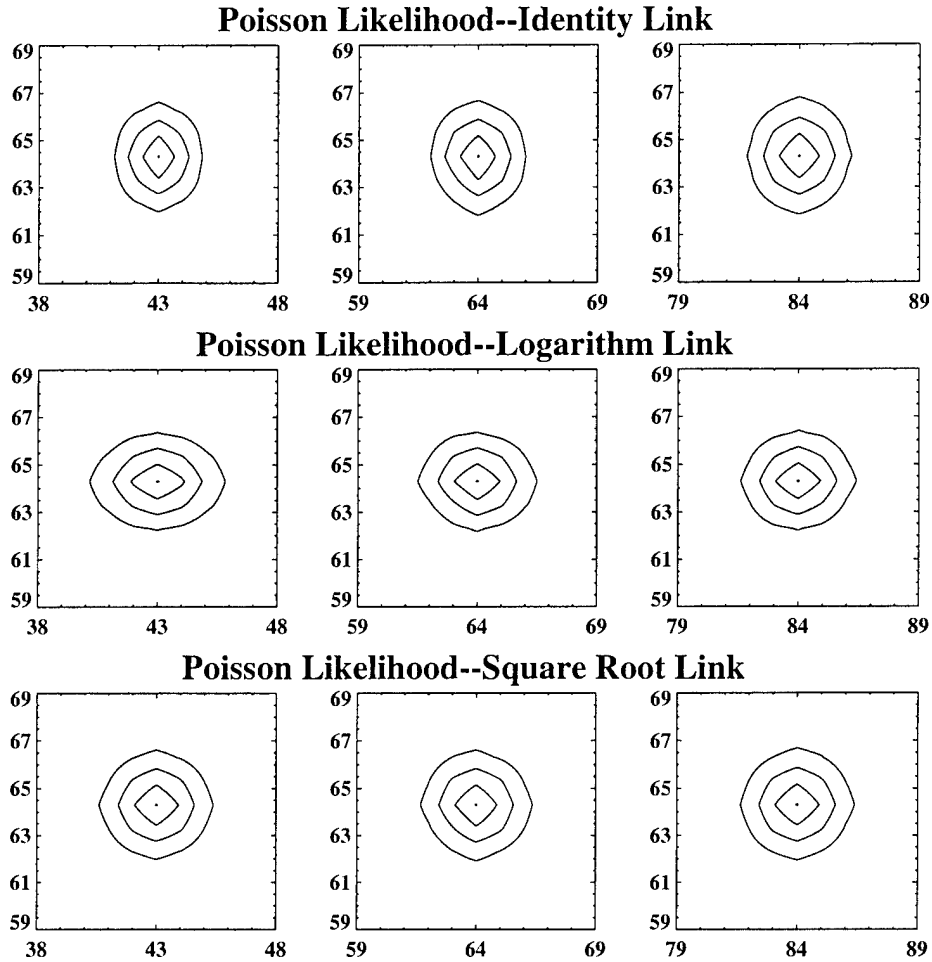
Fig. 4. Contours of the LIR functions at 25, 50 ,75, and 99% of peak value. The left column corresponds to the center of the cold disk, the center column to the center of the background ellipse, and the right column to the center of the hot disk. The different link functions are seen to give rise to different resolution isotropy properties in addition to the different resolution uniformity properties already observed in Fig. 2.
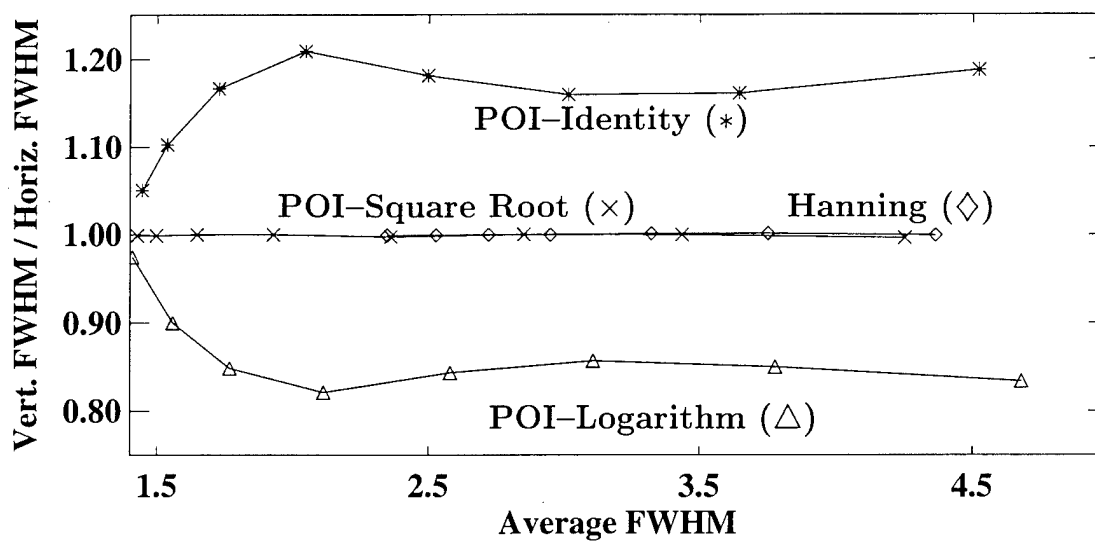
Fig. 5. Ratio of the vertical LIR function FWHM to the horizontal LIR function FWHM for the center of the background ellipse, plotted versus average FWHM for the center of the background ellipse. This figure confirms the anisotropic resolution properties observed in Fig. 4.
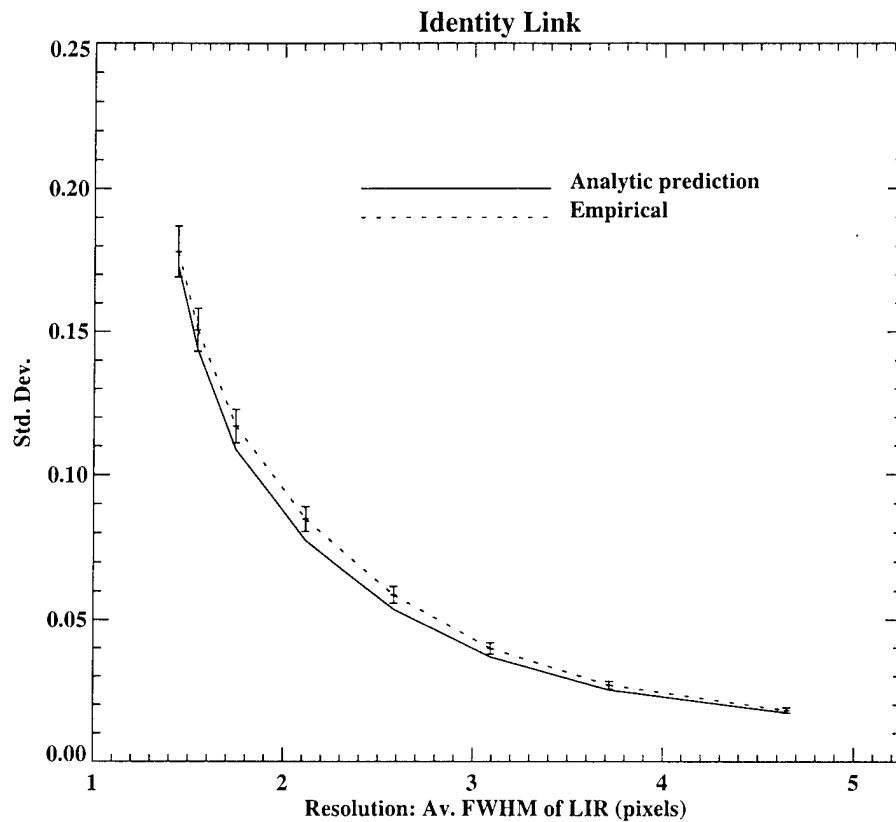
Identity Link

Fig. 6. Plot of analytic prediction of standard deviation at the center of the hot disk for the Poisson likelihood estimator with an identity link along with the empirical estimate of the same quantity, including 95% confidence intervals. The failure of the analytic prediction to intersect the 95% confidence intervals at most points suggests that the locally linear approximation underlying the analytic prediction may be unwarranted.
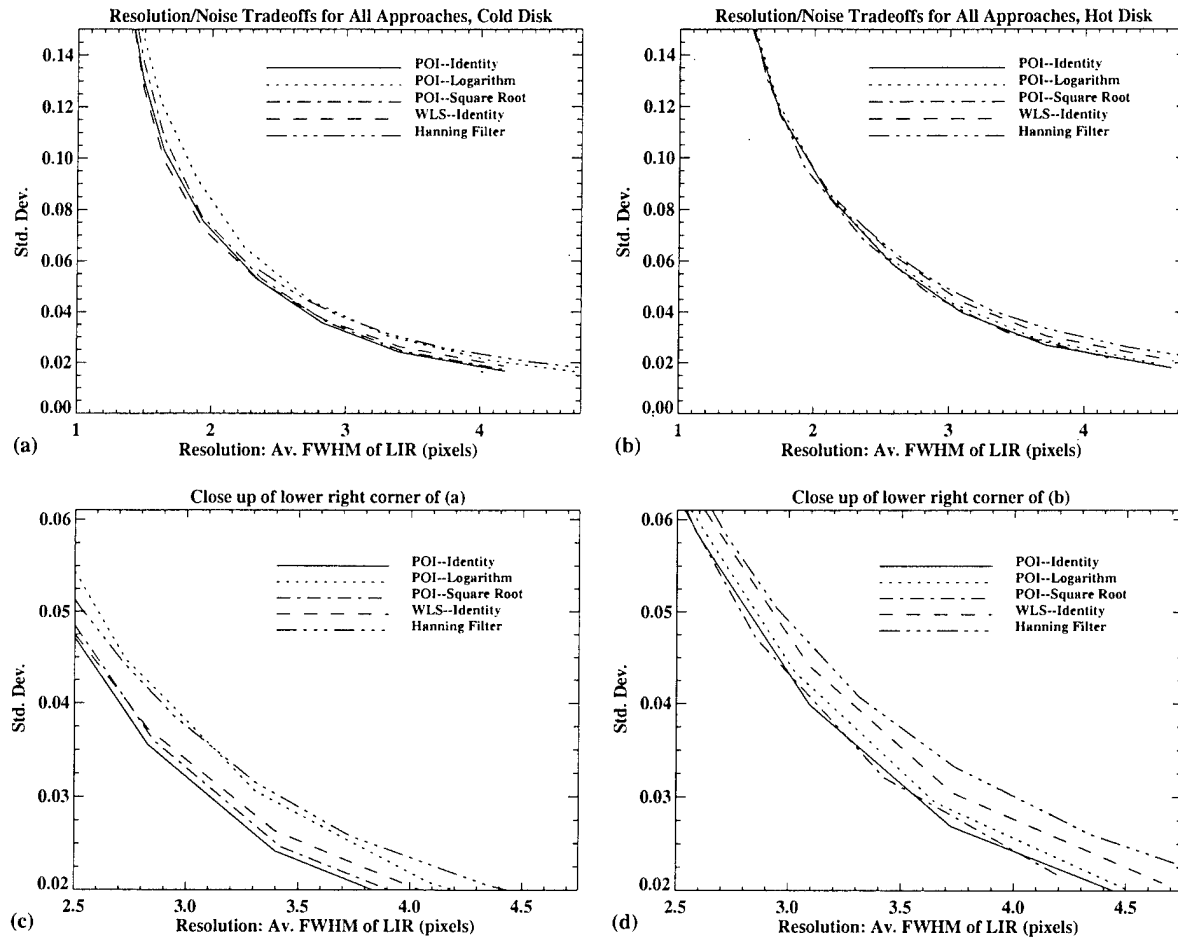
Fig. 7. (a) Resolution-variance tradeoff at the point at the center of the cold disk for the Poisson likelihood estimator with three choices of link function, the WLS estimator with the identity link, and the Hanning filter. (b) Same plot for the point at the center of the hot disk. (c) A close-up of the lower right corner of (a). (d) A close-up of the lower right corner of (b).
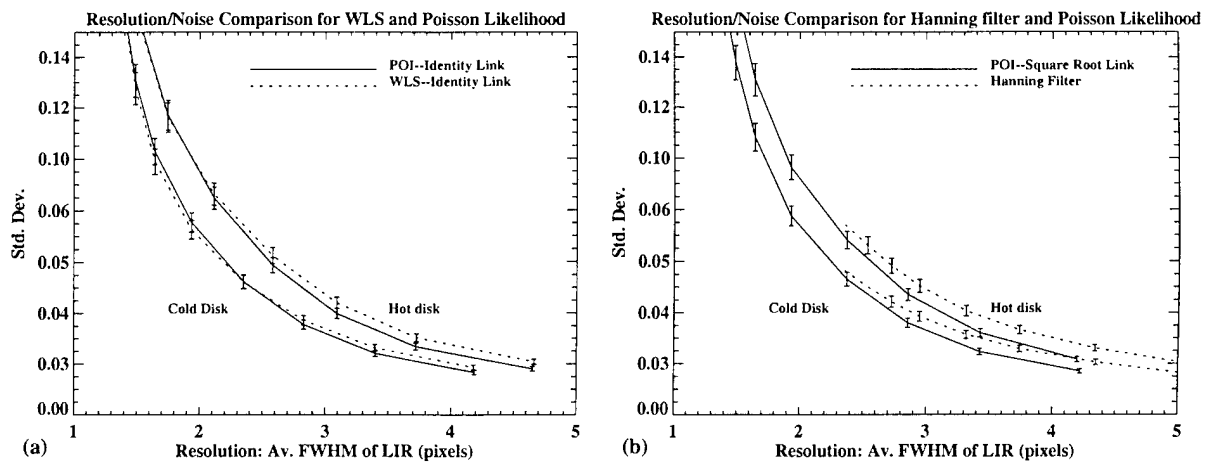
Fig. 8. (a) Resolution-variance tradeoff, with 95% confidence intervals, for the Poisson likelihood and WLS estimators, both with an identity link and thus matched resolution anisotropy, at the centers of the cold and hot disks. (b) Resolution-variance tradeoff, with 95% confidence intervals, for the Poisson likelihood estimator with a square root link and for the Hanning filter, both of which have essentially isotropic resolution, at the centers of the cold and hot disks.

Spline-Based Inverse Radon Transform in Two and Three Dimensions

P.J. La Rivière, *Student Member, IEEE,* and X. Pan, *Member, IEEE*

Department of Radiology, The University of Chicago, Chicago, IL, 60637

Abstract

While the exact inverse Radon transform is a continuous integral equation, the discrete nature of the data output by tomographic imaging systems generally demands that images be reconstructed using a discrete approximation to the transform. However, by fitting an analytic function to the projection data prior to reconstruction, one can avoid such approximations and preserve and exploit the continuous nature of the inverse transform.

We present methods for the evaluation of the inverse Radon transform in two and three dimensions in which cubic spline functions are fit to the projection data, allowing the integrals that represent the filtration of the sinogram to be carried out in closed form and also eliminating the need for interpolation upon backprojection. Moreover, in the presence of noise, the algorithm can be used to reconstruct directly from the coefficients of *smoothing* splines, which are the minimizers of a popular curve-fitting measure. We find that the 2D and 3D direct-spline algorithms have superior resolution to their 2D and 3D FBP counterparts, albeit with higher noise levels, and that they have slightly lower ideal-observer signal-to-noise ratios for the detection of a 1-cm, spherical lesion with a 6:1 lesion-background concentration ratio.

I. INTRODUCTION

The inverse Radon transform provides the mathematical foundation for tomographic imaging, which involves reconstructing images of distributions of anatomical or physiological properties from projections of those distributions. In computed tomography (CT) [1], for instance, the property being imaged is the linear photon attenuation coefficient of tissue at various points in the body, while in positron emission tomography (PET) [2] or in single-photon emission computed tomography (SPECT) [3] it is the concentration of injected radiopharmaceuticals at various points in the body. We will denote any such distribution in two dimensions by $f(x,y)$ and label each projection through it by the pair $\{\xi,\varphi\}$, where φ specifies the *projection angle* and ξ the *projection distance*. The value of such a projection is given by the line integral of the distribution along the line specified by $\{\xi,\varphi\}$ and will be denoted by $p(\xi,\varphi)$. Similarly, we will denote a three-dimensional distribution by $f(x,y,z)$ and label each projection through it by the triplet $\{\xi,\theta,\varphi\}$, where θ and φ specify the orientation of the plane and ξ the distance of the plane to the origin of the coordinates. The value of such a projection is given by the *planar* integral of the distribution over the plane specified by $\{\xi,\theta,\varphi\}$ and will be denoted by $p(\xi,\theta,\varphi)$. The functions $p(\xi,\varphi)$ and $p(\xi,\theta,\varphi)$ are known as *sinograms* because in the two-dimensional case a point

distribution in $\{x,y\}$ space maps to a sinusoid in $\{\xi,\varphi\}$ space.

The Radon transform is a continuous, integral transform that relates the sinogram $p(\xi,\varphi)$ to $f(x,y)$ in two dimensions and the sinogram $p(\xi,\theta,\varphi)$ to $f(x,y,z)$ in three dimensions [4,5,6]. Inverting the Radon transform exactly to recover a distribution requires continuous, noise-free knowledge of the distribution's sinogram, which entails having an infinite set of perfect projection measurements. In practice, of course, one can only collect projection data in the two-dimensional case for a finite number of projection distances ξ_i (we call these discrete samples *projection bins*) at a finite number of projection angles φ_j, and the measurements are invariably contaminated with noise. In the three-dimensional case, the planar-integral projection data cannot generally be measured directly and must instead be generated from line-integral projection data; it is, however, still only generated for a finite number of projection bins ξ_i and projection angles φ_j and θ_k.

Because the sinogram $p(\xi,\varphi)$ in two dimensions (or $p(\xi,\theta,\varphi)$ in three dimensions) is known only on a finite grid of ξ_i and φ_j (or ξ_i, φ_j, and θ_k), we cannot invert the Radon transform exactly to recover the distribution $f(x,y)$ (or $f(x,y,z)$); we must instead turn to a discrete approximation of the inverse. For instance, one way of implementing the most popular two-dimensional Radon inversion algorithm—filtered backprojection (FBP) [7,8]—begins with a discrete filtration of the sinogram. The filtered samples of ξ_i for each projection angle φ_j are then backprojected onto the image grid and the resulting images summed to give a final reconstructed image. One way to view the backprojection step is to imagine casting a perpendicular ray from each image pixel to each projection angle in turn, summing the sinogram values picked up at each angle to obtain the final pixel value. In this view, the difficulty lies in determining what value to pick up from each projection angle, for in general the perpendicular line will not fall directly in the center of a projection bin. In the simplest schemes, one simply picks up the value of the bin the pixel projects onto, while in a more complicated approach one might perform a linear interpolation of the values in the two nearest projection bins. In the most sophisticated schemes, one uses a weighted average of the values in a slightly larger neighborhood. A similar procedure can be used to implement three-dimensional FBP [9].

The discreteness of the sinogram thus dictates discreteness in both the filtration and backprojection steps of the algorithm. If, however, one had a continuous, analytic expression for the sinogram at each projection angle—if the sinogram were a set of known one-dimensional functions of ξ—it might be possible to implement the filtration and backprojection steps in a continuous manner. The filtration could be performed analytically, and the resulting filtered

projections would be continuous functions which, in the pixel-traversal view of backprojection described above, could be sampled wherever a projection might strike without any need to interpolate. Naturally, such an analytic, continuous expression for the sinogram cannot be obtained directly from any tomographic imaging system, but must rather be obtained by fitting an analytic expression to the discretely sampled data. Not every class of function that could be fit to the data would allow the filtration of the projections to be calculated in closed form, but Wahba [10] has shown that one class that does allow such a solution are the cubic splines, piecewise third-order polynomials that are continuous up to and including the second derivative at the connection points between pieces. This is the class of fitting functions we investigate in this paper, introducing Wahba's results (with some corrections and simplifications), and extending the treatment to the three-dimensional Radon transform.

This method offers a certain conceptual appeal, as well as the advantage of directness when one wishes to smooth noisy projection data by fitting curves that minimize the popular penalized least-squares measure [11]. As it turns out, the minimizers of this measure are the natural cubic splines mentioned above and thus reconstruction can proceed directly from the coefficients of these splines in this case.

II. METHODS

A. Inverse 2D Radon transform in coordinate space

The essential problem in two-dimensional tomography is the reconstruction of a distribution $f(x,y)$ (or $f(r,\theta)$ in polar coordinates) from knowledge of the discrete sinogram $p(\xi_i, \varphi_j)$, where $i = -N,...,N$ and $j = 1,...,M$. This convention for the index i, particularly the choice of an odd number of projection bins, will simplify the mathematical expressions to be derived below, but the proposed technique is applicable to geometries with an even number of bins as well. We assume in the present paper that we have a parallel-beam geometry.

Perhaps the most familiar way of expressing the inverse of the two-dimensional Radon transform is in terms of the frequency-space representation of the continuous sinogram:

$$f(r,\theta) = \int_0^\pi \int_{-\infty}^{\infty} |v| P(v,\varphi) e^{j2\pi v\xi} dv \, d\varphi, \quad (1)$$

where v is the spatial-frequency variable corresponding to ξ, $P(v,\varphi)$ is the Fourier transform of the sinogram $p(\xi,\varphi)$ with respect to the variable ξ, and j is the imaginary number $\sqrt{-1}$. This expression provides the theoretical basis for FBP, as $|v|$ is just the familiar ramp filter. This expression may be written in coordinate space as

$$f(r,\theta) = \frac{1}{2\pi^2} \int_0^\pi J_{r,\theta}(\varphi) d\varphi, \quad (2)$$

where

$$J_{r,\theta}(\varphi) = \left\{ \lim_{\varepsilon \to 0} \left[\int_{-\infty}^{\xi'-\varepsilon} \frac{p'(\xi,\varphi)}{\xi'-\xi} d\xi + \int_{\xi'+\varepsilon}^{\infty} \frac{p'(\xi,\varphi)}{\xi'-\xi} d\xi \right] \right\}, \quad (3)$$

and in which $\xi' = r\cos(\theta - \varphi)$ and $p'(\xi,\varphi)$ is the first derivative of the sinogram $p(\xi,\varphi)$ with respect to ξ [12,4]. Taking the limit as $\varepsilon \to 0$ of the sum of these two integrals allows us to avoid integrating over the singularity at $\xi' = \xi$. In general equations (2) and (3) are less useful than equation (1) because the data collected in PET, SPECT, or CT constitute samples of the sinogram $p(\xi,\varphi)$ itself, and do not provide any direct information about $p'(\xi,\varphi)$. However, by fitting an analytic, differentiable function of ξ to the projection data at each angle, we could obtain an expression for $p'(\xi,\varphi)$. If $p'(\xi,\varphi)$ had an auspicious functional form, we would then be able to solve the integrals in equation (3) in closed form.

B. Interpolating and smoothing splines

To obtain an expression for the sinogram that is analytic and differentiable with respect to the variable ξ, we fit a function to the projection data at each angle. That is, for each angle φ_j, we fit a one-dimensional function of ξ, $\hat{p}_{\varphi_j}(\xi)$, to the sinogram values $p_{\varphi_j}(\xi_i)$ measured on the $2N+1$ abscissas ξ_i $(i = -N,...,N)$. If the data is noiseless, it is desirable to use a function that passes through the points $p_{\varphi_j}(\xi_i)$, which we call an *interpolating* curve, while if the data is noisy a *smoothing* curve may be more appropriate. One fitting framework that can handle both of these situations is known as penalized least-squares [11], in which the function $\hat{p}_{\varphi_j}(\xi)$ is chosen to be the minimizer of the functional

$$F\left[\hat{p}_{\varphi_j}(\xi)\right] = \sum_{i=-N}^{N} \left[p_{\varphi_j}(\xi_i) - \hat{p}_{\varphi_j}(\xi_i) \right]^2 + \lambda \int_a^b \left(\hat{p}''_{\varphi_j}(\xi) \right)^2 d\xi, \quad (4)$$

where a and b are the endpoints of the interval on which the curve $\hat{p}_{\varphi_j}(\xi)$ is to be defined. The first term in equation (4) is the familiar squared-error measure, while the second is a measure of the smoothness of the fit curve. The parameter λ thus mediates the tradeoff between the competing goals of achieving a good fit to the data and maintaining a smooth curve. By choosing λ to be zero, we eliminate the smoothness constraint and ensure that the minimizing curve will be an interpolant to the data; if λ grows large, the smoothness constraint dominates and the curve approaches a linear fit to the data. For intermediate values of λ the minimizing curve balances the goodness-of-fit and smoothness constraints. The parameter λ can be chosen *a priori* [13] or it can be determined from the statistics of the data using an automatic procedure such as generalized cross-validation (GCV) [14].

The minimizers of this functional F belong to the class of functions known as natural cubic splines [13]. These are piecewise polynomial curves that join at the abscissa values ξ_i, where they are continuous up to and including the second derivative. They can be represented as

$$\hat{p}_{\varphi_j}(\xi) = a_i + b_i\xi + c_i \xi^2/2 + d_i \xi^3/3, \quad \xi \in [\xi_i, \xi_{i+1}], \quad (5)$$

where a_i, b_i, c_i, and d_i are constants that fully specify the spline curve on the interval $[\xi_i, \xi_{i+1}]$. Of interest to the Radon inversion problem is the fact that we can approximate the first derivative of the sinogram for fixed angle φ_j by

$$p'(\xi,\varphi_j) \cong \hat{p}'_{\varphi_j}(\xi) = b_i + c_i\xi + d_i\xi^2, \quad \xi \in [\xi_i, \xi_{i+1}]. \quad (6)$$

C. Spline-based inverse of the 2D Radon transform

Given this analytic expression for the derivative of the sinogram, we can proceed with the inversion of the two-dimensional Radon transform in equation (2). While the dimensional Radon transform in equation (2). While the sinogram now has a continuous representation in the variable ξ, it is still discrete in the angular variable. Assuming that the M angular samples are equally spaced over 180° or 360° (the result is the same in either case), the integral in equation (2) can be approximated by the sum

$$f(r,\theta) \cong \frac{1}{2\pi M} \sum_{j=1}^{M} J_{r,\theta}(\varphi_j), \qquad (7)$$

where $J_{r,\theta}(\varphi_j)$ is given by equation (3). For a given coordinate (r,θ) in image space, and for a given projection angle φ_j, $\xi' = r\cos(\theta - \varphi_j)$. We label the projection bin that ξ' falls in by m, that is, $\xi' \in [\xi_m, \xi_{m+1}]$. Using the expression for $p'(\xi,\varphi)$ given by equation (6), $J_{r,\theta}(\varphi_j)$ can be expressed as

$$J_{r,\theta}(\varphi_j) = \sum_{\substack{i=-N \\ i \neq m}}^{N-1} \int_{\xi_i}^{\xi_{i+1}} \frac{b_i + c_i \xi + d_i \xi^2}{\xi' - \xi} d\xi$$

$$+ \lim_{\epsilon \to 0}\left[\int_{\xi_m}^{\xi'-\epsilon} \frac{b_m + c_m \xi + d_m \xi^2}{\xi' - \xi} d\xi + \int_{\xi'+\epsilon}^{\xi_{m+1}} \frac{b_m + c_m \xi + d_m \xi^2}{\xi' - \xi} d\xi \right], \qquad (8)$$

where the integrals of equation (3) are now expressed as sums of integrals over the subintervals between the original abscissa points, with appropriate accommodation for the singularity at ξ'. These integrals can be solved in closed form [10], and the resulting expression contains some potentially unstable terms. However, by combining these terms in a particular way and invoking spline identities, a stable form can be derived. The details are given in appendix A, where the final expression for $J_{r,\theta}(\varphi_j)$ is shown to be

$$J_{r,\theta}(\varphi_j) = T + \sum_{\substack{i=-N \\ i \neq m-1,m,m+1}}^{N-1} (b_i + c_i \xi' + d_i \xi'^2) \ln\left(\frac{\xi' - \xi_i}{\xi' - \xi_{i+1}} \right)$$

$$- \sum_{i=-N}^{N-1} (c_i + 2d_i \xi')(\xi_{i+1} - \xi_i) + \frac{1}{2} \sum_{i=-N}^{N-1} d_i \left[(\xi' - \xi_i)^2 - (\xi' - \xi_{i+1})^2 \right] \qquad (9)$$

where T is given by equation (A.4) of appendix A.

D. The 3D Radon transform in coordinate space

The essential problem in three-dimensional computed tomography is the reconstruction of a distribution $f(x,y,z)$ from knowledge of the discrete planar-integral sinogram $p(\xi_i, \theta_k, \varphi_j)$, where $i = -N,...,N$, $j = 1,...,M_\varphi$, and $k = 1,...,M_\theta$ [15]. In general, these planar integrals are not measured directly by tomographic imaging systems, but must rather be calculated by "rebinning" the line integrals that are measured directly [9].

The inverse three-dimensional Radon transform has a form similar to the two-dimensional case, with a few differences that greatly simplify the task of evaluating it numerically. Specifically,

$$f(x,y,z) = \frac{-1}{4\pi^2} \int_0^\pi \int_0^\pi p''(\xi',\theta,\varphi) \sin\theta \, d\varphi \, d\theta, \qquad (10)$$

where

$$\xi' = x\sin\theta\cos\varphi + y\sin\theta\sin\varphi + z\cos\theta, \qquad (11)$$

and $p''(\xi',\theta,\varphi)$ is the second derivative of the three-dimensional sinogram with respect to ξ'. This expression differs in two principal ways from the expression for the two-dimensional inverse Radon transform given by equations (2) and (3). First, it now involves the second derivative of the sinogram with respect to ξ' rather than the first derivative and second, the convolution in ξ has disappeared. This reflects the fact that an inverse Radon transform of odd degree can be calculated using purely *local* information—the value of the image at a point (x,y,z) can be determined solely from information at points in the sinogram space that (x,y,z) projects onto, rather than from a convolution integral over all points in sinogram space as in the even-dimensional case [5,16].

E. Spline-based inverse of the 3D Radon transform

As in the two-dimensional case, we wish to fit an analytic function of ξ to each sequence of ξ_i labeled by a distinct pair $\{\varphi_j, \theta_k\}$. We do so using the spline formalism described above and obtain

$$\hat{p}_{\varphi_j,\theta_k}(\xi) = a_i + b_i \xi + c_i \xi^2/2 + d_i \xi^3/3, \quad \xi \in [\xi_i, \xi_{i+1}]. \quad (12)$$

Consequently the second derivative of the sinogram is approximately

$$p''(\xi,\theta_k,\varphi_j) \cong \hat{p}''_{\varphi_j,\theta_k}(\xi) = c_i + 2d_i \xi, \quad \xi \in [\xi_i, \xi_{i+1}]. \quad (13)$$

Now as in the two-dimensional case, the discreteness of the angular samples means that the two integrals give way to sums and we write

$$f(x,y,z) \cong -\frac{1}{4} \frac{1}{M_\theta} \frac{1}{M_\varphi} \sum_{j=1}^{M_\varphi} \sum_{k=1}^{M_\theta} J_{x,y,z}(\theta_k,\varphi_j)\sin\theta_k, \qquad (14)$$

where

$$J_{x,y,z}(\theta_k,\varphi_j) = c_i + 2d_i \xi', \quad \xi' \in [\xi_i, \xi_{i+1}], \qquad (15)$$

and

$$\xi' = x\sin\theta_k\cos\varphi_j + y\sin\theta_k\sin\varphi_j + z\cos\theta_k. \qquad (16)$$

The simplicity of the three-dimensional inversion is now apparent, for the functions J can be evaluated in a straightforward manner whereas in the two-dimensional case evaluation of the functions J involved performing a complicated integral over ξ and taking care in handling numerically unstable terms.

F. Application to phantom and real data

In order to demonstrate that the 2D direct-spline inverse of the Radon transform produces images of comparable or superior quality to filtered backprojection (FBP), we reconstructed images of a numerical Hoffman brain phantom [17] using both methods. The sinogram consisted of 128 simulated noiseless projections of the phantom, taken over 360° and each comprising 400 projection bins. The sinogram contained a total of 1.72x10⁸ counts. We first reconstructed the phantom using standard area-weighted FBP with a ramp filter (cutoff=1.0 times the Nyquist frequency). We then fit an

interpolating spline to the data at each of the 128 projection angles and used the 2D direct-spline technique to reconstruct.

Poisson noise was then added to the sinogram. Images were reconstructed using FBP with a ramp filter (cutoff=1.0) as well as the direct-spline technique using an interpolating spline in order to see how the algorithms compared in the presence of noise without prior smoothing. Smoothing splines were then fit to the projection data at each angle, using GCV to determine the smoothing parameter, and an image reconstructed from the coefficients using the direct-spline method. In order to examine the performance of FBP in the face of data with the same degree of smoothness, we sampled these smoothing splines to generate a discrete sinogram that was reconstructed using FBP with a ramp filter (cutoff=1.0).

For the 3D case, we reconstructed images of a Data Spectrum ventricular phantom from projection data acquired on a Picker XP3000 three-headed SPECT system fitted with high-resolution, parallel-hole collimators. The phantom was filled with 3.27 mCi of Tc-99m and placed at the center of rotation. Each head collected data on a 128x128 grid and at 120 projection angles over 360°. We rebinned the projection data from a single head to generate planar integrals on a 128x60x120 grid. Images were reconstructed from this planar-integral data using 3D FBP, and also using the 3D direct-spline inversion method after splines were fit to the data as described in section II.E. We then fit smoothing splines to the planar-integral data and reconstructed directly from the spline coefficients. Finally, we sampled the smoothing splines to generate a smoothed, discrete sinogram and used that as input to the 3D FBP algorithm.

G. Resolution, noise, and signal-to-noise studies

G.1. Resolution

In order to compare quantitatively the resolutions of the direct-spline algorithms (both two- and three-dimensional) with their FBP counterparts, we acquired projection images of a small (1 cm) spherical lesion containing 7.6 mCi of Tc-99m placed in an 800 cc cylindrical phantom containing cold (zero-activity) water. A Picker XP2000 two-headed SPECT system fitted with ultra-high-resolution, parallel-hole collimators was used. The heads rotated at their minimum radius of rotation (9 cm) and acquired 120 views over 360° onto a 128x128 matrix (pixel size=4.67 mm).

For the two-dimensional algorithms, we extracted the 2D sinogram corresponding to the slice through the center of the lesion and reconstructed images using FBP with a ramp filter (cutoff=1.0) as well as using the direct spline inversion with interpolating splines. The reconstructed lesion was approximately a symmetric 2D Gaussian in shape and we determined its full-width half-maximum by collapsing it into a one-dimensional function and fitting this profile with a Gaussian curve. For the three-dimensional reconstruction algorithms, we rebinned the projection data to generate planar-integral data on a 128x60x120 grid. We then used 3D FBP and the 3D direct spline method (using interpolating splines) to reconstruct the slice through the center of the lesion and determined the FWHM of the resulting Gaussian by the same method as above.

In order to isolate the contribution of the reconstruction algorithm to the FWHM of the lesion in the reconstructed images, the contribution from the lesion's inherent width as well as the imaging system's point-spread function had to be removed. The net effect of these two factors was estimated by determining the average FWHM of the lesion as it appeared in the 120 projection images. Assuming then that the reconstruction algorithms could be characterized by Gaussian point-spread functions, the FWHM of these functions were determined by subtracting (in quadrature) the average projection FWHM from the FWHMs of the reconstructed lesions discussed above.

G.2. Noise Levels

To characterize the noise level in images reconstructed by the direct-spline methods and their FBP counterparts, we acquired 20 1-minute projection datasets of the same 800 cc cylindrical phantom used in the previous section, this time containing 3.7 mCi of Tc-99m and no lesion. One slice of this uniform cylinder was reconstructed for each of the 20 datasets using 2D FBP, 3D FBP, 2D spline inversion, and 3D spline inversion (both of these using interpolating splines). For a given algorithm, the same six circular regions of interest (ROI) were examined in each of the 20 slices and the coefficient of variation (the standard deviation of the pixel values in the ROI divided by the mean of the pixel values in the ROI) calculated for each of the 120 ROIs. The average of these 120 coefficients of variation was then computed.

G.3. Signal-to-noise ratio

It is not uncommon for a reconstruction algorithm to offer enhanced resolution at the price of amplified noise. The overall effect of such a tradeoff is sometimes better characterized by computing a signal-to-noise ratio (SNR). We used the two datasets described above to compute so-called ideal-observer SNRs. The ideal-observer framework [18] offers a way of assessing the amount of information the data output by an imaging device (possibly modified by image processing) contains with regard to the performance of a specified task. For linear imaging processes in which the noise in the output image is assumed to be additive, Gaussian, zero-mean, stationary, and independent of the presence or absence of the signal, the ideal-observer framework allows us to characterize fully the quality of the imaging system data with respect to the performance of the specified signal-detection task with a single number, the ideal observer SNR. This can be expressed as

$$SNR_i^2 = \int \frac{|\Delta S_{out}(\upsilon)|^2}{W(\upsilon)} d\upsilon, \tag{17}$$

where $|\Delta S_{out}(\upsilon)|^2$ is the power spectrum of the signal in output space, i.e. after it has been degraded by the imaging system and possibly image processing, and $W(\upsilon)$ is the Wiener spectrum.

In order to calculate the ideal-observer SNR of the reconstructed images, we regard the 20 reconstructed images of the cylinder alone described in section G.2 as an ensemble of so-called background images. By adding the projections of the lesion described in section G.1 with a suitable scaling factor (chosen in this case to produce a 6:1 lesion-background

concentration ratio) to the projections of the cylinder and then reconstructing, we can generate an ensemble of signal-plus-background images. Then $|\Delta S_{out}(\upsilon)|^2$ is easily determined by computing the power spectrum of the difference between the two ensemble averages. The Wiener spectrum $W(\upsilon)$ can be computed from the ensemble of background images. The detailed procedure for the calculation is as follows:

1. The lesion projections were scaled to simulate a desired lesion-background concentration ratio (6:1 in this case) and added to each of the 20 sets of background projections.

2. Images of the slice through the center of the lesion were reconstructed for the 20 signal-plus-background datasets using eight different methods: the 2D spline-based inversion with interpolating splines, 2D FBP with a ramp filter (cutoff=1.0), the 2D spline-based inversion using smoothing splines, 2D FBP using a sinogram sampled from these smoothing splines, the 3D spline-based inverse using interpolating splines, 3D FBP, the 3D spline-based inverse using smoothing splines, and 3D FBP using a sinogram sampled from these smoothing splines.

3. The 20 corresponding sinograms of background alone were reconstructed in the same eight ways.

4. An average signal image was determined for each reconstruction method by subtracting the average of the 20 background reconstructions from the average of the 20 signal-plus-background reconstructions. The signal power spectrum was computed by squaring the Fourier transform of this image.

5. While SPECT images are not stationary in general, the attenuated projections of a uniform cylinder of this diameter are quite flat over a broad central region, and thus one might reasonably expect the reconstructed images of this cylinder to be locally stationary near their center, precisely where the lesion is expected to lie [19]. A "local" Wiener spectrum in this region was thus computed from the 20 images of background alone by subtracting the average background image from each of the individual background images, resulting in 20 noise images. Each such image was multiplied by a circularly symmetric window of the form:

$$w(\mathbf{r}) = 1 \quad \text{for } |\mathbf{r}| \leq 0.9R,$$

$$w(\mathbf{r}) = 0.5 \times (1 + \cos(\pi(|\mathbf{r}| - 0.9R)/0.2R),$$
$$\text{for } 0.9R < |\mathbf{r}| < 1.1R, \text{ and}$$

$$w(\mathbf{r}) = 0 \quad \text{for } |\mathbf{r}| \geq 1.1R,$$

(19)

where R is the radius of the circular region over which the noise is expected to be stationary (chosen to be 6 pixels) and $\mathbf{r}$ the radial position in the image. The power spectrum of each of the 20 images was computed by taking the square of the Fourier transform of the resulting image. The 20 power spectra were then averaged and scaled so that the volume under the Wiener spectrum equaled the average variance in the circle of radius R [20-22].

6. The ideal-observer SNR was then determined by summing the quotient of the calculated signal spectrum and Wiener spectrum.

The ideal-observer SNR of the raw projection data represents an upper bound on the ideal-observer SNR of the reconstructed images, a point that is discussed further in section IV. Because the noise in the projection data is uncorrelated, the ideal-observer SNR can be computed using a simpler expression than equation (17),

$$SNR_I^2 = \Delta s' \left(diag\left\{ \left\langle p(\xi_i, \varphi_j) \right\rangle \right\} \right)^{-1} \Delta s, \quad (18)$$

where Δs is the signal projection in the spatial domain expressed as a one-dimensional vector, $diag\{\ \}$ denotes a diagonal matrix and $\left\langle p(\xi_i, \varphi_j) \right\rangle$ is the noise-free background projection data [23]. This is estimated by the sample mean of the 20 noisy background projection datasets.

Finally, it should be noted that in using the ideal-observer framework at all it is implicitly being assumed that the data satisfies the assumptions discussed above: that the system is linear and that the noise in the planar or reconstructed images is additive, Gaussian, stationary, zero-mean, and independent of the presence or absence of the signal. Given the reasonably high count levels (~10-15/pixel), the fact that the signal is relatively small and low contrast, and the discussions of stationarity above, these assumption about the noise are not unreasonable. The requirement of linearity seemingly undermines the use of the framework to analyze images that are reconstructed from smoothing splines that have been fit using an adaptive, and thus non-linear algorithm. However, what is truly required for equation (17) to be meaningful is not linearity in the face of any possible input but more specifically that the system transfer function be the same whether the particular signal of interest is present or absent from the particular background of interest. Again, because the signal in question is relatively small and low contrast, it should not greatly affect the noise properties of the projection images and thus the use of smoothing splines should yield a similar effective system transfer function whether the signal is present or absent.
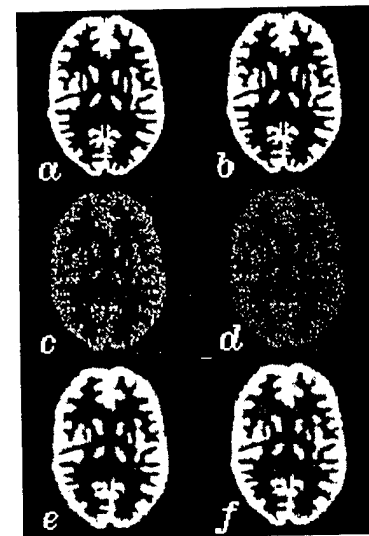

Figure 1. Reconstructions of a Hoffman brain phantom from simulated projections without noise (a-b) and with Poisson noise added (c-f). Reconstruction methods are: (a) FBP with ramp filter (cutoff=1.0), (b) direct-spline inversion using interpolating splines, (c) FBP with ramp filter (cutoff=1.0), (d) direct-spline inversion using interpolating splines, (e) FBP from a sinogram generated by sampling smoothing splines, (f) direct-spline inversion using smoothing splines.

III. RESULTS

The results of reconstructing the Hoffman brain phantom with and without noise using both the 2D direct-spline inverse and 2D FBP are depicted in Figure 1. The algorithms clearly yield qualitatively similar results.

The results of reconstructing the ventricular phantom data using 3D direct-spline inversion with interpolating splines, 3D FBP, 3D direct-spline inversion with smoothing splines, and 3D FBP using a sinogram resampled from the smoothing splines are depicted in Figure 2. The algorithms are again seen to yield qualitatively similar results.

The resolution measurements for the four basic algorithms—2D direct-spline inversion, 2D FBP with a ramp filter (cutoff=1.0), 3D direct-spline inversion, and 3D FBP— are summarized in Table 1. The results indicate that the direct-spline inversions have superior resolution to FBP in both the 2D and 3D cases and also that the 2D algorithms have superior resolution to the 3D algorithms.

The results of the noise study are summarized in Table 2, where it can be seen that the noise level in the direct-spline reconstructions is higher than that in the FBP reconstructions and that the noise level in the 2D reconstructions is higher than that in the 3D reconstructions.

Finally, the ideal-observer signal-to-noise ratio results are summarized in Table 3 for the four basic algorithms as well as their counterparts using smoothing splines. We observe that the direct-spline algorithms have slightly lower SNRs than their FBP counterparts when using interpolating splines, while the SNRs become comparable when using smoothing splines. Furthermore, the use of smoothing splines seems to have little effect on SNR in the 2D case while degrading it in the 3D case. All of the reconstructed images are seen to have lower SNRs than the raw projection data, a fact that is discussed in greater detail in the next section. Typical images reconstructed using each of these eight methods are shown in Figure 3.

Table 1
FWHM of in-plane reconstruction point-spread functions

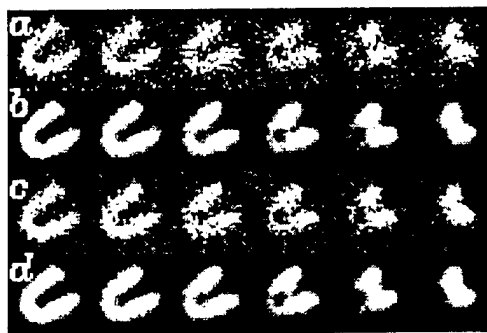| Algorithm	FWHM
2D direct spline	1.6 mm
2D FBP	4.5 mm
3D direct spline	3.9 mm
3D FBP	5.0 mm

Figure 2. Selected slices of a ventricular phantom reconstructed by (a) 3D direct-spline inversion using interpolating splines, (b) 3D direct-spline inversion using smoothing splines, (c) 3D FBP, and (d) 3D FBP from a sinogram obtained by sampling the smoothing splines in (b).

Table 2
Coefficients of variation for various reconstruction algorithms

Algorithm	COV
2D direct spline	0.60
2D FBP	0.39
3D direct spline	0.34
3D FBP	0.23

Table 3
Ideal-observer SNRs for various reconstruction algorithms

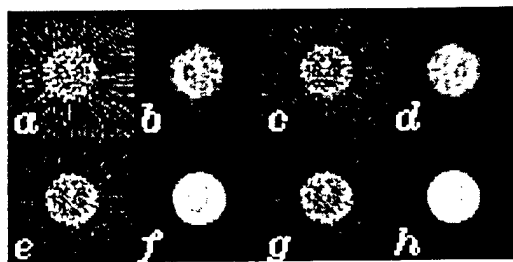| Algorithm	Ideal-observer SNR
Sinogram	10.7
2D interpolating spline	7.9
2D FBP	8.6
2D smoothing spline	8.4
2D FBP w/ smth. spline	8.4
3D interpolating spline	9.0
3D FBP	9.6
3D smoothing spline	7.5
3D FBP w/ smth. spline	7.0

Figure 3. Reconstructions of a selected slice of a cylindrical phantom containing a spherical lesion. Reconstruction methods: (a) 2D direct-spline inversion using interpolating splines, (b) 2D direct-spline inversion using smoothing splines, (c) 2D FBP, (d) 2D FBP from a sinogram obtained by sampling the smoothing splines in (b), (e) 3D direct-spline inversion using interpolating splines, (f) 3D FBP, (g) 3D direct-spline inversion using smoothing splines, (h) 3D FBP from a sinogram obtained by sampling the smoothing splines in (g)

IV. DISCUSSION AND CONCLUSIONS

As discussed in section I, the principal difference between the 2D and 3D direct-spline inversion algorithms and FBP is

in the nature of the interpolation step upon backprojection. The interpolation in FBP is simply less accurate than the more sophisticated cubic-spline interpolation used in the direct-spline method. The cruder FBP interpolation is more likely to smooth over high-frequency variations in the projection data than is cubic-spline interpolation, and thus it is not surprising that the FBP algorithms have inferior resolution to the spline-based algorithms, as illustrated in Table 1. However, because the high-frequency components of the data include considerable noise as well, the FBP algorithms would be expected to produce less noisy reconstructions than the spline-based reconstructions. This expectation is confirmed by the results of the noise study reported in Table 2.

The 3D direct-spline and FBP algorithms are both seen to have inferior resolution and lower noise levels than their 2D counterparts. This can be attributed to the fact that the 3D reconstruction process involves an additional averaging or smoothing step which occurs when the raw projection data is rebinned into a planar-integral sinogram by performing area-weighted forward projections of the 2D projection data at each projection angle.

Table 3 lists the ideal-observer SNR for the raw projection data and for the various reconstruction approaches. It is a fact that processing or even image reconstruction can never improve the ideal-observer SNR over that found in the raw projection data. However, these operations can certainly *diminish* the SNR if they are in some way singular and if the signal vector has components in the null space. We observe that all the reconstructed image SNRs are in fact lower than that of the raw projection data. The differences among those reconstructed image SNRs give some clue as to how much of the information contained in those projections is preserved in the reconstructed image. For instance, we see that ideal-observer SNRs are slightly lower for the 2D and 3D direct-spline inversions using interpolating splines than for their FBP counterparts using ramp filters. For this particular detection task, then, the interpolating-spline algorithm's amplification of noise outweighs the improvement it affords in resolution relative to FBP. However, when the noise is mitigated prior to reconstruction, as when the projection data has been fitted with smoothing splines, the SNR gap between the spline algorithms and FBP is considerably narrowed in the 2D case and reversed in the 3D case. This suggests that the spline-based algorithms may be of greatest use when resolution is paramount and the data contains relatively little noise, a situation more often encountered in computed tomography than in nuclear medicine.

The use of smoothing splines is seen to provide little or no improvement in SNR in the 2D case. Smoothing does affect the ideal-observer SNRs in the 3D case, but for the worse. It is clear from examining the images of Figure 3 that the reconstructed images using 3D algorithms and smoothing splines have an oversmoothed appearance. This can most likely be attributed to the fact, mentioned above, that the 3D reconstruction process effectively involves a prior smoothing during the rebinning step. The adaptive smoothing algorithm certainly makes allowances for the lower variability in the rebinned data and smooths this data less than it would the raw projection data. However, the modified statistics of the rebinned data simply do not agree as well with the statistical

model assumed by the smoothing algorithm, so it is perhaps not surprising that it yields a sub-optimal result. It remains a topic for further investigation as to whether smoothing the raw projection data prior to the rebinning step produces a better result.

V. ACKNOWLEDGMENTS

VI. REFERENCES

[1] G.N. Hounsfield, "Computerized transverse axial scanning (tomography). I. Description of system," *Br. J. Radiology*, vol. 46, pp. 1023-1047, 1973.

[2] M.M. Ter-Pegossian, M.E. Phelps, E.J. Hoffman and N.A. Mullani, "A positron-emission transaxial tomograph for nuclear imaging (PETT)," *Radiology*, vol. 114, pp. 89-98, 1975.

[3] T.F. Budinger, "Single-photon emission computed tomography," in *Diagnostic Nuclear Medicine*, M.P. Sandler *et al.*, eds., Baltimore: Williams and Wilkens, 1996, pp. 121-138.

[4] H.H. Barrett, "The Radon transform and its applications," in *Progress in Optics XXI*, E. Wolf, ed. Amsterdam: Elsevier Science Publishers, 1984, pp. 217-286.

[5] S. Deans, *The Radon Transform and Some of its Applications*, New York: Wiley, 1983.

[6] G.T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*, Boston: Academic Press, 1980.

[7] L.A. Shepp and B.F. Logan, "The Fourier reconstruction of a head section," *IEEE Trans. Nucl. Sci.*, vol. 21, pp. 21-43, 1974.

[8] R.H. Huesman, G.T. Gullberg, W.L. Greenberg and T.F. Budinger, *User Manual, Donner Algorithms for Reconstruction Tomography*, Lawrence Berkeley Laboratory, University of California, 1977.

[9] C. Wu, "Fully three-dimensional reconstruction in PET and SPECT by the use of three-dimensional Radon transforms," *PhD Thesis*, University of Chicago, 1994.

[10] G. Wahba, "A new approach to the numerical evaluation of the inverse Radon transform with discrete, noisy data," *Technion Preprint Series* MT-479, 1980.

[11] P.J. Green and B.W. Silverman, *Nonparametric Regression and Generalized Linear Models*, London: Chapman Hall, 1994.

[12] G.T. Herman G T and A. Naparstek, "Fast image reconstruction based on a Radon inversion formula

appropriate for rapidly collected data," *SIAM J. Applied Mathematics*, vol. 33, pp. 511-533, 1977.

[13] J.A. Fessler, "Tomographic reconstruction using information-weighted spline smoothing," in *Information Processing in Medical Imaging*, H.H. Barrett and A.F. Gmitro, eds., Berlin: Springer-Verlag, 1993, pp. 290-300.

[14] D.M. Bates, M.J. Lindstrom, G. Wahba and B. Yandell "GCVPACK routines for generalized cross-validation," *Commun. Statist. Simul. Comput.*, vol. 16, pp. 263-297, 1987.

[15] M.Y. Chiu, H.H. Barrett and R.G. Simpson, "Three-dimensional image reconstruction from projections," *J. Opt. Soc. Am.*, vol. 70, pp. 755-762, 1980.

[16] F. Natterer, *The Mathematics of Computerized Tomography*, Stuttgart: John Wiley & Sons, 1986.

[17] E.J. Hoffman, P.D. Cutler, W.M. Digby and J.C. Mazziotta, "3-D phantom to simulate cerebral blood flow and metabolic images for PET," *IEEE Trans. Nucl. Sci.*, vol. 37, pp. 616-620, 1990.

[18] "Medical imaging—the assessment of image quality," *ICRU Report 54*, 1996. Available from ICRU, 7910 Woodmont Ave., Bethesda, MD, 20814.

[19] S.C. Moore, M.F. Kijewski, S.P. Müller and B.L. Holman, "SPECT image noise power: effects of nonstationary projection noise and attenuation compensation," *JNM*, vol. 29, pp. 1704-1709, 1988.

[20] S.J. Riederer, N.J. Pelc and D.A. Chesler, "The noise power spectrum in computed x-ray tomography," *Phys. Med. Biol.*, vol. 23, pp. 446-454, 1978.

[21] K. Hanson, "Detectability in computed tomographic images," *Med. Phys.*, vol. 6, pp. 441-451, 1979.

[22] M.F. Kijewski and P.F. Judy, "The noise power spectrum of CT Images," *Phys. Med. Biol.*, vol. 32, pp. 565-575, 1987.

[23] J. Yao and H.H. Barrett, "Predicting human performance by a channelized Hotelling observer model," *Proc. SPIE*, vol. 1768, pp. 161-168, 1992.

APPENDIX A

This appendix supplies the details connecting equation (9) to equation (8). Solving the integrals in equation (8) yields

$$J_{r,\theta}(\varphi_j) = \sum_{\substack{i=-N \\ i\neq m}}^{N-1}\left(b_i + c_i\xi' + d_i\xi'^2\right)\ln\left(\frac{\xi' - \xi_i}{\xi' - \xi_{i+1}}\right) - \sum_{i=-N}^{N-1}\left(c_i + 2d_i\xi'\right)\left(\xi_{i+1} - \xi_i\right)$$
$$+ \frac{1}{2}\sum_{i=-N}^{N-1}d_i\left[\left(\xi' - \xi_i\right)^2 - \left(\xi' - \xi_{i+1}\right)^2\right] + \left(b_m + c_m\xi' + d_m\xi'^2\right)\ln\left(\frac{\xi' - \xi_m}{\xi_{m+1} - \xi'}\right).$$

(A.1)

There are three terms in equation (A.1), containing $\ln[(\xi' - \xi_{m-1})/(\xi' - \xi_m)]$, $\ln[(\xi' - \xi_{m+1})/(\xi' - \xi_{m+2})]$, and $\ln[(\xi' - \xi_m)/(\xi_{m+1} - \xi')]$, respectively, which are numerically unstable when ξ' is near ξ_m in the case of the first term, near ξ_{m+1} in the case of the second, or near either in the case of the third. The three terms, whose sum we denote as T are

$$T = \left(b_{m-1} + c_{m-1}\xi' + d_{m-1}\xi'^2\right)\ln\left(\frac{\xi' - \xi_{m-1}}{\xi' - \xi_m}\right) + \left(b_{m+1} + c_{m+1}\xi' + d_{m+1}\xi'^2\right)\ln\left(\frac{\xi' - \xi_{m+1}}{\xi' - \xi_{m+2}}\right) + \left(b_m + c_m\xi' + d_m\xi'^2\right)\ln\left(\frac{\xi' - \xi_m}{\xi_{m+1} - \xi'}\right).$$

(A.2)

Spline identities that follow from the fact that a cubic spline has continuous first and second derivatives at ξ_m,

$$\left(b_m - b_{m-1}\right) + \left(c_m - c_{m-1}\right)\xi_m + \left(d_m - d_{m-1}\right)\xi_m^2 = 0$$

(A.3)

$$\left(c_m - c_{m-1}\right) + 2\left(d_m - d_{m-1}\right)\xi_m = 0,$$

and a similar pair reflecting the continuity at ξ_{m+1} can be used to remove the singularities in equation (A.2). Using these identities we can reexpress T as

$$T = \left(b_{m-1} + c_{m-1}\xi' + d_{m-1}\xi'^2\right)\ln\left(\xi' - \xi_{m-1}\right) + \left(d_m - d_{m-1}\right)\left(\xi' - \xi_m\right)^2\ln\left(\xi' - \xi_m\right)$$

(A.4)

$$+ \left(d_{m+1} - d_m\right)\left(\xi_{m+1} - \xi'\right)^2\ln\left(\xi_{m+1} - \xi'\right) - \left(b_{m+1} + c_{m+1}\xi' + d_{m+1}\xi'^2\right)\ln\left(\xi_{m+2} - \xi'\right),$$

in which the second and third terms pose no numerical instability because $\left(\xi' - \xi_m\right)^2\ln\left(\xi' - \xi_m\right) \rightarrow 0$ as $\xi' \rightarrow \xi_m$ and $\left(\xi_{m+1} - \xi'\right)^2\ln\left(\xi_{m+1} - \xi'\right) \rightarrow 0$ as $\xi' \rightarrow \xi_{m+1}$.

Hence the final expression for $J_{r,\theta}(\varphi_j)$ is given by

$$J_{r,\theta}(\varphi_j) = \sum_{\substack{i=-N \\ i\neq m-1,m,m+1}}^{N-1}\left(b_i + c_i\xi' + d_i\xi'^2\right)\ln\left(\frac{\xi' - \xi_i}{\xi' - \xi_{i+1}}\right) - \sum_{i=-N}^{N-1}\left(c_i + 2d_i\xi'\right)\left(\xi_{i+1} - \xi_i\right) + \frac{1}{2}\sum_{i=-N}^{N-1}d_i\left[\left(\xi' - \xi_i\right)^2 - \left(\xi' - \xi_{i+1}\right)^2\right] + T,$$

(A.5)

where T is given by equation (A.4).

Ideal-Observer Analysis of Lesion Detectability in Planar, Conventional SPECT, and Dedicated SPECT Scintimammography Using Effective Multi-Dimensional Smoothing

P.J. La Rivière, *Student Member, IEEE*, X. Pan, *Member, IEEE*, and B.C. Penney

Department of Radiology, The University of Chicago, Chicago, IL, 60637

Abstract

Scintimammography, a nuclear-medicine imaging technique that relies on the preferential uptake of Tc-99m-sestamibi and other radionuclides in breast malignancies, has the potential to provide differentiation of mammographically suspicious lesions, as well as outright detection of malignancies in women with radiographically dense breasts. In this work we use the ideal-observer framework to quantify the detectability of a 1-cm lesion using three different imaging geometries: the planar technique that is the current clinical standard, conventional single-photon emission computed tomography (SPECT), in which the scintillation cameras rotate around the entire torso, and dedicated breast SPECT, in which the cameras rotate around the breast alone. We also introduce an adaptive smoothing technique for the processing of planar images and of sinograms that exploits Fourier transforms to achieve effective multidimensional smoothing at a reasonable computational cost.

For the detection of a 1-cm lesion with a clinically typical 6:1 tumor-background ratio, we find ideal-observer signal-to-noise ratios (SNR) that suggest that the dedicated breast SPECT geometry is the most effective of the three, and that the adaptive, two-dimensional smoothing technique should enhance lesion detectability in the tomographic reconstructions.

I. INTRODUCTION

Breast cancer is the most frequently diagnosed invasive malignancy among American women and ranks second only to lung cancer in annual cancer-related mortality for this group [1]. Numerous studies have shown that early detection and treatment of breast cancer can improve survival rates [2-4]. Screen-film mammography has come to play a vital role in this detection process, due to its high (80-90%) sensitivity to breast malignancies. However, mammography is notoriously poor at distinguishing benign from malignant tumors, having reported specificities and positive predictive values of 15-30% [5]. This means that only a small percentage of lesions biopsied on the basis of suspicious mammographic appearance are found to be malignant.

In recent years, researchers have developed and studied a nuclear-medicine test with the potential to provide relatively low-cost, minimally invasive differentiation of breast abnormalities identified by physical examination or mammography [6-18]. Known as scintimammography, the test relies on the preferential uptake of Tc-99m-sestamibi or other radionuclides such as Tl-201, Tc-99m- tetrofosmin, or Tc-99m-MDP in breast malignancies as compared to normal breast tissue or benign abnormalities. Indeed, one study has shown that typical *in vivo* tumor-background concentration ratios of Tc-99m-sestamibi are 5.64±3.06 [19]. This focal uptake can be imaged in a number of ways, though the most widely used clinical protocol involves acquiring one or two planar views—one lateral view and possibly an additional oblique or anterior view—while the patient lies prone on a specially designed table [20]. The imaging time is typically 10-15 minutes per view. Numerous clinical studies with histological follow-up have been performed using this or a similar protocol, reporting sensitivities of 83-96%, and specificities of 66-100% when using Tc-99m-sestamibi to image mammographically suspicious lesions [8-18]. In addition to differentiating breast abnormalities detected by other means, scintimammography may also have a role in the detection of breast malignancies in patients with radiographically dense breasts, for whom screen-film mammograms are often difficult to interpret [7].

A few of these studies of scintimammography have also examined the role of conventional SPECT (where the patient lies supine and the camera circles the torso) in detecting focal uptake of Tc-99m-sestamibi and have found comparable but not generally improved sensitivities as compared to planar techniques, along with substantially lower specificities [14,21-23]. Wang *et al.* [24] speculated that this surprisingly poor performance was due to substantial attenuation of photons emitted in the breast by the torso in at least half of the views as well as to the presence of scatter from organs such as the heart and liver known to have high uptake of Tc-99m-sestamibi. The poor performance of conventional SPECT may also be related to the inferior resolution of conventional SPECT as compared to planar techniques in this situation, due to the fact that the scintillation cameras are on average further away from the breast in the conventional SPECT geometry than in the planar geometry. Wang *et al.* also investigated a geometry that they called vertical axis-of-rotation SPECT and that we call dedicated breast SPECT, or simply dedicated SPECT, in which the scintillation cameras are assumed to rotate around one breast alone. This geometry eliminates the effect of attenuation by the thorax and, with proper shielding, the effect of scatter from the thorax. Moreover, the small radius of rotation offers improved resolution and sensitivity. In phantom studies, Wang *et al.* found that with this dedicated geometry they were able to detect a breast lesion with an outer diameter of 1 cm and a 6:1 lesion-to-background concentration ratio that was not detectable in either conventional SPECT or planar studies with the same total imaging time.

The present work examines quantitatively the question of lesion detectability in these three different geometries—planar, conventional SPECT, and dedicated SPECT—using the so-called ideal-observer framework to calculate signal-to-noise ratios as a function of lesion concentration for the different geometries, and, in the case of the tomographic geometries, for different reconstruction filters. It also

introduces and applies an adaptive smoothing technique for the processing of planar images or sinograms that exploits Fourier transforms to achieve effective multidimensional smoothing at a reasonable computational cost. Such pre-processing is found to improve idealized lesion detectability in the reconstructed images.

II. METHODS

A. Effective multi-dimensional smoothing

The projection data acquired by tomographic nuclear-medicine imaging devices is invariably contaminated by noise, which is propagated and often amplified by the reconstruction algorithm. Filtered backprojection (FBP), the most computationally efficient and most commonly used of these algorithms, attempts to control this noise by incorporating an apodization window, such as a Hanning filter, into the projection filtration step. Alternatively, in some reconstructions, only the mathematically exact ramp filter is applied to the projection data, but the reconstructed image is subjected to a two- (or three-) dimensional post-filtering. While somewhat different in application and effect, these two approaches have the common property of being linear; the degree of smoothing applied to the data does not depend on the data, but is rather fixed *a priori*, and this same degree of smoothing is applied uniformly to the entire dataset.

In contrast, non-linear, adaptive smoothing methods, such as the generalized cross-validation method to be discussed below, determine the degree of smoothing to be applied to various specified subsets of the data from the statistics of each such subset and the degree of smoothing can vary from subset to subset. For instance, in 2D image reconstruction, the data at each projection angle could be smoothed differently, with the degree of smoothing determined from the statistics of the data at each angle.

When smoothing for 2D image reconstruction, one would in principle like to exploit the statistical correlations between different projections as well as those within a given projection. A truly 2D adaptive smoothing operation of this type can be computationally expensive and difficult to implement [25]. However, by exploiting the properties of the Fourier transform, one can achieve *effective* 2D smoothing at the cost of a series of 1D smoothing operations. Consider a 2D discrete sinogram $p(\xi_i, \theta_j)$, where ξ_i refers to the ith projection bin $(i=1,...,N)$ and θ_j the jth projection angle $(j=1,...,M)$. It can be shown [26] that the following sequence of operations is equivalent to an adaptive, 2D smoothing of the sinogram:

1. Take a 1D discrete Fourier transform of the sinogram with respect to the projection angle θ_j; the result can be viewed as a set of 1D functions of the untransformed variable ξ_i, each labeled by an angular frequency index k, i.e. $P_k(\xi_i)$.

2. Perform an adaptive 1D smoothing of each of these M functions of ξ_i, yielding M discrete smoothed functions $P_k^s(\xi_i)$, where the superscript s stands for smoothed.

3. Perform an inverse 1D discrete Fourier transform of $P_k^s(\xi_i)$ with respect to the angular frequency k to recover $p^s(\xi_i, \theta_j)$.

The adaptive 1D smoothing we perform on each of the M functions $P_k(\xi_i)$ is known as penalized least-squares smoothing [27,28], and involves fitting the discrete data with a continuous smoothing curve $P_k^s(\xi)$ that minimizes the functional

$$S\{P_k^s(\xi)\} = \sum_{i=1}^{N}\left[P_k(\xi_i) - P_k^s(\xi_i)\right]^2 + \alpha\int_0^T [(P_k^s(\xi)'')^2]\,d\xi.$$

where ξ is a continuous variable representing the position along a given projection, T is the total length of the projection, and the double prime denotes the second derivative with respect to ξ. The two terms in this functional represent the competing goals of achieving a good fit to the data while maintaining a smooth curve, with the parameter α mediating the tradeoff. For instance, if α is zero, the smoothness constraint disappears and the minimizing curve will be a piecewise linear interpolant to the data; if α grows large, the smoothness constraint dominates and the curve approaches a simple linear fit to the data. For intermediate values of α, the minimizing curves balance the goodness-of-fit and smoothness constraints. It can in fact be shown that the minimizers of this functional will always be members of a class of functions known as natural cubic splines [27,28]. These are piecewise cubic curves that join at the abscissa values ξ_i, where they are continuous up to and including the second derivative.

Clearly the choice of α determines the degree of smoothing that is applied to the data, and it is this parameter that is determined from the statistics of the data itself in an *adaptive* implementation of penalized least-squares smoothing using an algorithm known as generalized cross-validation [29]. Thus a generally different α is used in the smoothing of each of the M functions $P_k(\xi_i)$. The resulting continuous smoothed functions $P_k^s(\xi)$ must then be sampled to yield the discrete functions $P_k^s(\xi_i)$ which are subjected to the inverse DFT in step 3 above.

It should also be noted that while we have, for the sake of simplicity, discussed effective 2D smoothing, the technique can be extended to any number of dimensions. To smooth an n-dimensional function, one can simply take an $(n-1)$-dimensional Fourier transform of the function and then perform a set of 1D smoothings over the untransformed variable prior to taking an inverse $(n-1)$-dimensional Fourier transform [26].

B. Ideal Observer Framework

The ideal-observer framework [30] offers a way of assessing the amount of information the data from an imaging device contain with regard to the performance of a specified task. For example, the simplest such task is the detection of a signal of known strength, shape, and location in a specified background. In this case, the framework seeks to quantify the degree to which an ideal observer—one who can use the information contained in the images to its fullest extent—can reliably distinguish images containing the

background alone from images containing the background and the signal when both kinds of images are corrupted by noise, blurring, and other imperfections. For linear imaging processes in which the noise in the output image is assumed to be additive, Gaussian, zero-mean, stationary, and independent of the presence or absence of the signal, the ideal-observer framework allows us to characterize fully the quality of the imaging system data with respect to the performance of the specified signal-detection task by a single number, the ideal-observer signal-to-noise ratio (SNR). This is usually expressed as

$$SNR_i^2 = K^2 \int \frac{\left|\Delta S_{in}(\upsilon)\right|^2 MTF^2(\upsilon)}{W(\upsilon)}d\upsilon, \qquad (2)$$

where K is the large-scale transfer characteristic of the imaging system at the desired operating point, $MTF(\upsilon)$ is the system modulation transfer function, $W(\upsilon)$ is the system Wiener spectrum and $\left|\Delta S_{in}(\upsilon)\right|^2$ is the power spectrum of the signal in input space, i.e. before it has been scaled and degraded by the imaging system. This expression allows one to determine the ideal-observer SNR for any analytically specified input signal once K, $MTF(\upsilon)$, and $W(\upsilon)$ are known. Alternatively, if one wishes to determine the ideal-observer SNR for a particular real signal, equation (2) could be re-expressed as

$$SNR_i^2 = \int \frac{\left|\Delta S_{out}(\upsilon)\right|^2}{W(\upsilon)}d\upsilon, \qquad (3)$$

where $\left|\Delta S_{out}(\upsilon)\right|^2$ is the power spectrum of the signal in output space, i.e. after it has been scaled and degraded by the imaging system. This is the form we will use, because by acquiring an ensemble of images containing the signal and the background as well as an ensemble of images containing the background alone, $\left|\Delta S_{out}(\upsilon)\right|^2$ can be easily determined by computing the power spectrum of the difference between the two ensemble averages.

Finally, when the noise in an image is uncorrelated, the ideal-observer SNR takes on a particularly simple form,

$$SNR_i^2 = \Delta s^t \left(diag\{Hf\}\right)^{-1} \Delta s, \qquad (4)$$

where Δs is the signal vector, $diag\{\ \}$ is a diagonal matrix and Hf is the noise free projection data [31].

C. Data Acquisition and Processing

As discussed above, computing the ideal-observer SNR for a given imaging geometry and processing approach requires two ensembles of images: one set consisting of images of the signal and background together and one set consisting of images of the background alone. In order to preserve the flexibility to compute the ideal-observer SNR for lesions of varying concentration, we acquired high-count projection images of the signal alone (i.e., in a cold background), which could be scaled as desired and added to the ensemble of background projections alone to produce an ensemble of signal-plus-background projections. These were then either analyzed directly for the planar geometry or processed and reconstructed for the tomographic geometries. Of course, for linear techniques, we could have computed $\left|\Delta S_{out}(\upsilon)\right|^2$ from the images of the signal alone, because the

difference between the ensemble averages in this case is simply equal to the ensemble average of the differences—the signal we added to the background ensemble. However, this equivalence does not generally hold for non-linear techniques such as the adaptive smoothing being investigated.

Our phantom consisted of an 14-cm diameter, 800 cc cylinder with a 1-cm outer diameter spherical lesion insert. This lesion size is representative of the smallest currently detected in scintimammography. For each geometry we acquired 20 images of the background alone, for which we put 3.7 mCi of activity in the phantom and imaged for 1 minute total for each conventional or dedicated SPECT acquisition and 30 seconds for each planar view. These combinations of activity and imaging time were chosen to produce clinically realistic count levels using the following reasoning. As with Wang et al., we began with the assumption that 1% of a typical 25 mCi clinical dose of Tc-99m-sestamibi is taken up by the myocardium. Using the volume of the myocardium in the Data Spectrum Corporation cardiac insert as a guide, along with the assumption that soft tissue will have a 1:15 concentration relative to the myocardium allowed us to determine the expected concentration of activity in healthy breast tissue. We wished to compare detectability in these three geometries given the same total imaging time, and assumed that typical clinical imaging times would be 30 minutes per SPECT study and 15 minutes per planar view. Thus we were comparing the three geometries for equal total imaging times given that two-view planar studies are common. Given this, we scaled up the calculated concentration by a factor of 30 and scaled down the imaging times by the same factor to allow for more rapid data acquisition. All imaging times were adjusted to compensate for the decay of the activity. To image the lesion we filled it with 7.6 mCi of Tc-99m, placed it in the cylinder now filled with cold (zero-activity) water, and imaged for 30 minutes in conventional and dedicated SPECT and 20 minutes for the planar view. This combination of activity and imaging time was chosen simply to provide high-count, low noise images of the signal, which could be scaled appropriately and added to the background images.

The dedicated SPECT images were acquired by placing the phantom at the center of rotation of a Picker XP2000 two-headed SPECT system with the heads rotating at the minimum radius of rotation (9.0 cm). In this configuration the heads were within 2.0 cm of the walls of the phantom. The breast phantom was not attached to an anthropomorphic torso phantom because Wang et al. showed that with proper shielding the contribution of scatter from the torso can be made negligible. We acquired 120 views over 360° with each head acquiring to a 128x128 matrix (pixel size=4.67 mm). We used a low-energy, ultra-high resolution collimator. The conventional SPECT images were also acquired in the absence of an anthropomorphic torso phantom, although the radius of rotation (25 cm) and the placement of the breast phantom (17 cm off-center) were determined with the torso phantom in place. The reason for this curious arrangement was to isolate the effect of the large radius of rotation on lesion detectability, without the additional degradations caused by attenuation or scatter in the torso. The ideal-observer SNR results for this arrangement will thus represent

an upper bound on the true detectability. This arrangement also facilitates computation of the Wiener spectrum, which requires images of a stationary, uniform background that would have been difficult to achieve in the presence of the highly non-uniform attenuation caused by the thorax. In other respects, the conventional SPECT images used the same acquisition parameters as the dedicated SPECT. Finally, the planar views were acquired with the phantom flush against one head, which acquired on a 128x128 matrix with a magnification factor of 2.0 (pixel size=2.33 mm).

For the tomographic geometries, ideal-observer SNRs for the sinograms were calculated using equation (4), while SNRs for reconstructed images were computed by the following procedure:

1. The signal projections were scaled to simulate a desired tumor-background concentration ratio (6:1 in this case) and added to each of the 20 sets of background projections.

2. The slice through the center of the lesion was selected and the 20 corresponding signal-plus-background sinograms were reconstructed by filtered backprojection using ramp and Hanning filters with various cutoff frequencies (0.4, 0.6, 0.8, and 1.0 times the Nyquist frequency). The sinograms were also processed using the effective 2D smoothing procedure described above and reconstructed by filtered backprojection.

3. The 20 corresponding sinograms of background alone were processed in the same way.

4. An average signal image was determined by subtracting the average of the 20 background alone reconstructions from the average of the 20 signal-plus-background reconstructions. The signal power spectrum was computed by squaring the Fourier transform of this image.

5. While SPECT images are not stationary in general, the attenuated projections of a uniform cylinder of this diameter are quite flat over a broad central region, and thus one might reasonably expect the reconstructed images of this cylinder to be locally stationary near their center, precisely where the lesion is expected to lie [32]. The "local" Wiener spectrum in this region was computed from the 20 images of background alone by subtracting the average background image from each of the individual background images, resulting in 20 noise images. Each such image was multiplied by a circularly symmetric window of the form:

$$w(\mathbf{r}) = 1 \quad \text{for } |\mathbf{r}| \le 0.9R,$$

$$w(\mathbf{r}) = 0.5 \times (1 + \cos(\pi(|\mathbf{r}| - 0.9R)/0.2R),$$
$$\text{for } 0.9R < |\mathbf{r}| < 1.1R, \text{ and}$$

$$w(\mathbf{r}) = 0 \quad \text{for } |\mathbf{r}| \ge 1.1R,$$

(5)

(with appropriate shifting for the off-center cylinder in the conventional geometry), where R is the radius of the circular region over which the noise is expected to be stationary (chosen to 6 pixels) and $\mathbf{r}$ the radial position in the image. The result was centered and zero-padded to 128x128. The noise power spectrum of each of the 20 images was computed by taking the square of the Fourier transform of the resulting image. The 20 noise power spectra were then averaged and scaled so that the volume under the Wiener spectrum equaled the average variance in the circle of radius R [33-35].

6. The ideal observer SNR was then determined by summing the quotient of the calculated signal spectrum and Wiener spectrum.

For planar views, the procedure was essentially the same, without the reconstruction step but including the application of the effective 2D smoothing. Signal spectra were determined from the difference between the ensemble averages of signal-plus-background images and background images and Wiener spectra from the background images alone, using the same rolled-off cylindrical window to isolate a reasonably uniform region and to minimize truncation effects. As a consistency check, equation (4) was also used to calculate the SNR of the unsmoothed planar dataset.

Finally it should be noted that in using the ideal-observer framework at all it is implicitly being assumed that the data satisfy the assumptions discussed in section II.B: that the system is linear and that the noise in the planar or reconstructed images is additive, Gaussian, zero-mean, stationary, and independent of the presence or absence of the signal. Given the reasonably high count levels (~10-15/pixel) the fact that the signal is relatively small and low contrast, and the discussion of stationarity in point 5 above, the assumptions about the noise seem reasonable. The requirement of linearity seemingly undermines the use of the framework to analyze images that have been processed by adaptive, effective multi-dimensional smoothing. However, what is truly required for equation (3) to be meaningful is not linearity in the face of any possible input but more specifically that the system transfer function be the same whether the particular signal of interest is present or absent from the particular background of interest. Again, because the signal in question is relatively small and low contrast, it should not greatly affect the noise properties of the projection images and thus the effective multi-dimensional smoothing algorithm should yield a similar effective system transfer function whether or not the signal is present.

III. RESULTS

The unsmoothed and smoothed planar views are shown in Figure 1. The ideal-observer SNR for the planar views was found to be 6.2 without processing (use of equations (3) and (4) yielded the same result) and essentially the same, 6.4, after adaptive, effective multi-dimensional smoothing
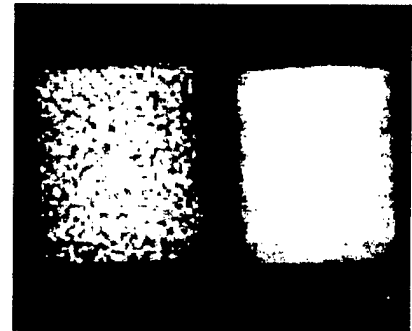

Figure 1: Planar images of a cylindrical phantom containing a 1-cm, 6:1 lesion. The image on the left is unprocessed; the image on the right has undergone adaptive, effective 2D smoothing. The image on the left corresponds to a calculated SNR of 6.2, the image on the right to an SNR of 6.4.

Table 1
Ideal Observer SNRs for conventional and dedicated SPECT

Processing Method	Dedicated Breast SPECT	Conventional SPECT[†]
Sinogram	10.7	6.7
Hanning (cutoff=0.4)	6.2	2.4
Hanning (cutoff=0.6)	8.0	3.1
Hanning (cutoff=0.8)	9.4	3.5
Hanning (cutoff=1.0)	9.9	3.7
Ramp (cutoff=0.4)	8.1	3.3
Ramp (cutoff=0.6)	9.7	3.7
Ramp (cutoff=0.8)	9.9	3.6
Ramp (cutoff=1.0)	9.3	3.7
Eff. 2D smoothing	10.6	5.5

[†]The conventional SPECT SNRs represent an upper bound on detectability (see section II.C for details).

The ideal-observer SNRs for the detection of a lesion with a 6:1 lesion-background concentration ratio are listed in Table 1 for the two different tomographic geometries.

It has been shown experimentally that the minimum SNR necessary for a human observer to be able to detect reliably a signal in a noisy background is 5.0 [36]. Regardless of the processing approach, the conventional SPECT and planar geometries yield SNRs that are below or only just above this threshold. The results above thus confirm quantitatively the findings of Wang et al. that a 6:1 lesion of this size is difficult or impossible to detect using planar or conventional SPECT geometries, but quite reliably detectable using a dedicated geometry. The results also indicate that effective two-dimensional smoothing provides improvement in SNR over other filtering approaches. Images corresponding to the two different geometries for selected processing methods are shown in Figure 2. These confirm visually the conclusions just stated: the 6:1 lesion is quite visible in all of the dedicated reconstructions, while it is effectively undetectable in the conventional reconstructions. The 6:1 lesion is rather difficult to discern in both the unprocessed and processed planar views depicted in Figure 1.

IV. DISCUSSION AND CONCLUSIONS

The SNR values given in Table 1 suggest that a dedicated SPECT geometry would lead to improved detectability for clinically typical lesions over the planar and conventional SPECT geometries. Recall that in the presence of attenuation and scatter from the torso we would expect the difference between the dedicated and conventional SPECT geometries to be even greater than it is here. The success of the dedicated geometry can be attributed to the fact that it combines the advantages of the other two approaches: because of its small radius of rotation, it offers good sensitivity and resolution comparable to that of a planar view acquired with the

scintillation camera flush against the phantom, while it offers the improved contrast offered by a tomographic system's ability to separate lesion activity from overlying and underlying activity.
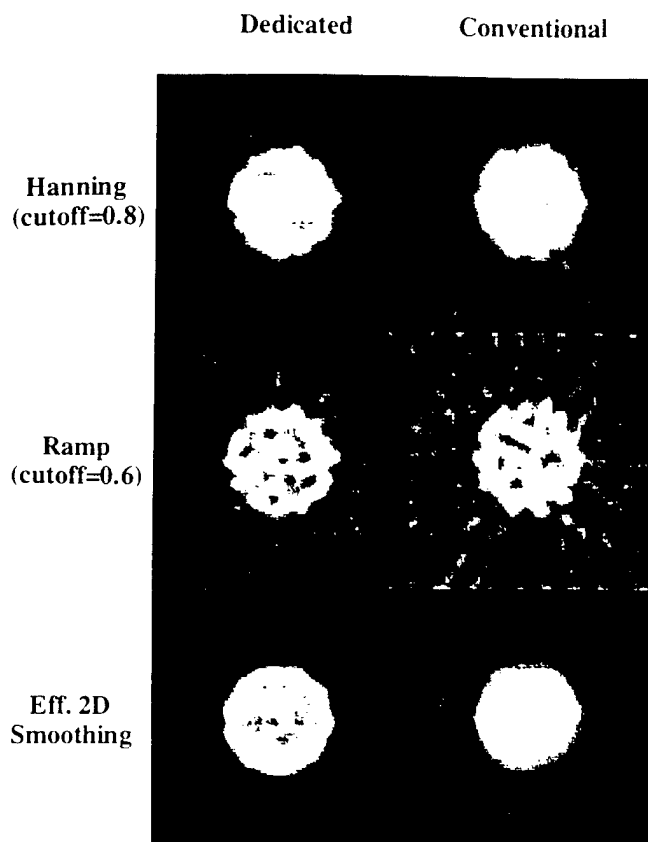

Figure 2: Reconstructed slices of a cylindrical phantom containing a 1-cm, 6:1 lesion for two different tomographic geometries.

The SNR values also support the hypothesis that adaptive, effective multi-dimensional smoothing may improve lesion detectability relative to the other kinds of filtering used in tomographic reconstruction. It is a fact that processing or even image reconstruction can never improve the ideal-observer SNR over that found in the raw projection data. However, these operations can certainly *diminish* the SNR if they are in some way singular and if the signal vector has components in the null space. We observe that all the reconstructed image SNRs are lower than those of their corresponding sinograms. The differences among those reconstructed image SNRs give some clue as to how much of the information contained in those projections is preserved in the reconstructed image. It would seem that the use of the effective multi-dimensional smoothing prior to reconstruction allows more of the information in the projections to persist in the reconstructed images.

Finally, it should be noted that in the case of fully linear imaging process, the calculated ideal-observer SNR should be directly proportional to the lesion concentration and thus to the tumor-background concentration ratio. This allows us to express the SNRs as a function of tumor-background

1278

concentration ratio, *C*, yielding for instance an SNR of 1.7*C for the dedicated SPECT geometry using a Hanning filter (cutoff=1.0), an SNR of 0.6*C for the conventional SPECT geometry using the same filter, and an SNR of 1.0*C for the unprocessed planar geometry. Assuming that an SNR of 5.0 corresponds to the threshold of detectability, this allows us to conclude that for the specified imaging times and lesion size, the minimum tumor-background ratios required for detectability are approximately 2.9-to-1 for the dedicated geometry, 8.3-to-1 for the conventional geometry, and 5.0-to-1 for the planar geometry using the specified filters.

Throughout this study we have been comparing the three geometries on the assumption of equal total imaging times (30 minutes) and implicitly assuming that in the planar geometry we were examining the one of the two 15-minute views in which the lesion appeared most prominently. It is not necessarily possible to identify this view *a priori*, but assuming it were we can estimate the SNR corresponding to a 30-minute acquisition at that single view. Because SNRs for linear methods should in principle scale as the square root of the imaging time, the SNR would be approximately 6.2 * $\sqrt{2}$ = 8.8, or comparable to the dedicated SPECT SNR.

The dependence of the ideal-observer SNR on lesion size remains a subject for future investigation.

V. ACKNOWLEDGMENTS

VI. REFERENCES

[1] Cancer Facts & Figures, Atlanta: American Cancer Society, 1996.

[2] I. Andersson, "Mammographic screening and mortality from breast cancer: Malmö mammographic screening trial," *Br. J. Med.*, vol. 297, pp. 943-948, 1988.

[3] J. Frisell, G. Eklund, L. Hellström, E. Lidbrink, L.E. Rutqvist and A. Somell, "Randomized study of mammographic screening—preliminary report on mortality in the Stockholm trial," *Breast Cancer Res. Treat.*, vol. 18, pp. 49-56, 1991.

[4] A.B. Miller, C.J. Baines, T. To and C. Wall, "Canadian national breast screening study," *Can. Med. Assoc. J.*, vol. 147, pp. 1459-1476, 1992.

[5] D.B. Kopans, "The positive predictive value of mammography," *AJR*, vol. 158, pp. 521-526, 1992.

[6] E. Bombardieri, F. Crippa, L. Maffioli and M. Greco, "Nuclear medicine techniques for the study of breast cancer," *Eur. J. Nucl. Med.*, vol. 24, pp. 809-824, 1997.

[7] A.D. Waxman, "The role of Tc-99m Methoxy-isobutylisonitrile in imaging breast cancer," *Seminars in Nuclear Medicine*, vol. 27, pp. 40-54, 1997.

[8] C.H. Kao, S.J. Wang and T.J. Liu, "The use of technetium-99m methoxyisobutylisonitrile breast scintigraphy to evaluate palpable breast masses," *Eur. J. Nucl. Med.*, vol. 21, pp. 432-436, 1994.

[9] I. Khalkhali, J. Cutrone, I. Mena, L. Diggles, R. Venegas, H. Vargas, B. Jackson and S. Klein, "Technetium-99m-sestamibi Scintimammography of Breast Lesions: Clinical and Pathological Follow-up," *J. Nucl. Med.*, vol. 36, pp. 1784-1789, 1995.

[10] I. Khalkhali, I. Mena and L. Diggles, "Review of imaging techniques for the diagnosis of breast canacer: a new role of prone scintimammography using technetium-99m-sestamibi," *Eur. J. Nucl. Med.*, vol., 21, pp. 357-362, 1994.

[11] I. Khalkhali, J. Cutrone, I. Mena, L. Diggles, R.J. Venegas, H.I. Vargas, B.L. Jackson, S. Khalkhali, J.F. Moss and S.R. Klein, "Scintimammography: the complementary role of Tc-99m sestamibi prone breast imaging for the diagnosis of breast carcinoma," *Radiology*, vol. 196, pp. 421-426, 1995.

[12] J. Villanueva-Meyer, M.H. Leonard Jr., E. Briscoe, F. Cesani, S.A. Ali, S. Rhoden, M. Hove and D. Cowan, "Mammoscintigraphy with Technetium-99m-sestamibi in suspected breast cancer," *J. Nucl. Med.*, vol. 37, pp. 926-930, 1996.

[13] R. Taillefer, A. Robidoux, R. Lambert, S. Turpin and Jean Laperrière, "Technetium-99m-sestamibi prone scintimammography to detect primary breast cancer and axillary lymph node involvement," *J. Nucl. Med.*, vol. 36, pp. 1758-1765.

[14] H. Palmedo, A. Schomburg, F. Grünwald, P. Mallman, D. Krebs and H.-J. Biersack, "Technetium-99m-MIBI scintimammography for suspicious breast lesions," *J. Nucl. Med.*, vol. 37, pp. 626-630, 1996.

[15] H. Palmedo, F. Grünwald, H. Bender, A. Schomburg, P. Mallmann, D. Krebs and H.J. Biersack, "Scintimammography with technetium-99m methoxy-isobutylisonitrile: comparison with mammography and magnetic resonance imaging," *Eur. J. Nucl. Med*, vol. 23, pp. 940-946, 1996.

[16] F. Scopinaro, M. Ierardi, L.M. Porfiri, N.S. Tiberio, G. De Vincentis, S. Mezi, P. Cannas, T. Gigliotti and L. Marzetti, "Tc-99m-MIBI prone scintimammography in patients with high and intermediate risk mammography," *Anticancer Research*, vol. 17, pp. 1635-1638, 1997.

[17] F. Scopinaro, O. Schillaci, W. Ussof, K. Nordling, R. Capoferro, G. De Vincentis, R. Danieli, M. Ierardi, V. Picardi, R. Tavolaro and A.C. Colella, "A three-center study on the diagnostic accuracy of 99mTc-MIBI scintimammography," *Anticancer Reaserch*, vol. 17, pp. 1631-1634, 1997.

[18] A. Maurier, D.F. Caroline, F.J. Jadali, T.A. Manzone, W.P. Maier, F.C. Au, and S.F. Schnall, "Limitations of caraniocaudal thallium-201 and technetium-99m-sestamibi mammoscintigraphy," *J. Nucl. Med.*, vol. 36, pp. 1696-1700, 1995.

[19] J. Maublant, M. de Latour, D. Mestas, A. Clemenson, S. Charrier, V. Feillel, G. Le Bouedec, P. Kaufmann, J. Dauplat and A. Veyre, "Technetium-99m-sestamibi uptake in breast tumor and associated lymph nodes," *J. Nucl. Med.*, vol. 37, pp. 922-925, 1996.

[20] L. Diggles, I. Mena and I. Khalkhali, "Technical aspects of prone dependent-breast scintimammography," *J. Nucl. Med. Tech.*, vol. 22, pp. 165-170, 1994.

[21] M.A. Nathan, J.E. Seabold, T. Barloon, J. Quesenberry, D.L. Bushnell, R.A. Robinson and D.C. Young, "Planar vs. SPECT tc-99m MIBI evaluation of suspicious breast lesions on mammograohy: histologic correlation," *J. Nucl. Med.*, vol. 35, p. 229P, 1994.

[22] N. Nagaraj, A. Waxman, G. Ashok, S. Khan, L. Memsic, J. Yadegar and E. Phillips, "Comparison of SPECT and planar Tc-99m sestamibi (MIBI) imaging in patients with carcinoma of the breast," *J. Nucl. Med.*, vol. 35, p. 299P, 1994.

[23] R. Tiling, M. Pechmann, H. Sommer, R. Moser, K. Tatsch and K. Khan, "Does SPECT improve the diagnostic accuracy of planar scintimammography with sestamibi?" *J. Nucl. Med.*, vol. 23, pp. 97-102, 1982.

[24] H. Wang, C. Scarfone, K.L. Greer, R.E. Coleman and R.J. Jaszczak, "Prone breast tumor imaging using vertical axis-of-rotation SPECT systems: an intital study," *IEEE Trans. Nucl. Sci.*, vol. 44, pp. 1271-1276, 1997.

[25] A. Williams and K. Burrage, "Surface fitting using GCV smoothing splines on supercomputers," *CD-ROM proceedings of the 1995 ACM/IEEE conference on supercomputing*, 1995. The proceedings are also available on the World Wide Web at *http://www.supercomp.org/sc95/proceedings.*

[26] X. Pan, "A general approach for multidimensional smoothing," *Med. Phys.* (in press), 1998.

[27] P.J. Green and B.W. Silverman, *Nonparametric Regression and Generalized Linear Models*, London: Chapman Hall, pp. 11-27, 1994.

[28] J.A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 13, pp. 290-300, 1994.

[29] D.M. Bates, M.J. Lindstrom, G. Wahba and B.S. Yandell, "GCVPACK routines for generalized cross validation," *Commun. Statist. Simul. Comput.*, vol. 16, pp. 263-297, 1987.

[30] "Medical imaging—the assessment of image quality," *ICRU Report 54*, 1996. Available from International Commission on Radiation Units and Measurements, 7910 Woodmont Ave., Bethesda, MD, 20814.

[31] J. Yao and H.H. Barrett, "Predicting human performance by a channelized Hotelling observer model," *Proc. SPIE*, vol. 1768, pp. 161-168, 1992.

[32] S.C. Moore, M.F. Kijewski, S.P. Müller and B.L. Holman, "SPECT image noise power: effects of nonstationary projection noise and attenuation compensation," *JNM*, vol. 29, pp. 1704-1709, 1988.

[33] S.J. Riederer, N.J. Pelc and D.A. Chesler, "The noise power spectrum in computed x-ray tomography," *Phys. Med. Biol.*, vol. 23, pp. 446-454, 1978.

[34] K. Hanson, "Detectability in computed tomographic images," *Med. Phys.*, vol. 6, pp. 441-451, 1979.

[35] M.F. Kijewski and P.F. Judy, "The noise power spectrum of CT Images," *Phys. Med. Biol.*, vol. 32, pp. 565-575, 1987.

[36] A. Rose, *Vision: Human and Electronic*, New York: Plenum Press, 1973.